



Gestion de risque de portefeuille : estimation de la VaR et CVaR

Mohamed Taha Chikhaoui

► To cite this version:

Mohamed Taha Chikhaoui. Gestion de risque de portefeuille : estimation de la VaR et CVaR. Probabilités [math.PR]. 2015. hal-01246153

HAL Id: hal-01246153

<https://inria.hal.science/hal-01246153>

Submitted on 18 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE SUPÉRIEURE PRIVÉE D'INGÉNIERIE ET DE TECHNOLOGIE
UNIVERSITÉ NICE SOPHIA ANTIPOLIS
INRIA



Septembre 2015
Soutenance du Projet de Fin d'Étude

Gestion de risque de portefeuille : Estimation de VaR et CVaR

Supervisé par : Mme Madalina DEACONU
M. Antoine LEJAY

Réalisé par : Maroua CHIKHAOUI

Remerciement

Au terme de ce travail, je tiens à remercier et à exprimer ma profonde gratitude et respectueuse reconnaissance à mes honorables encadrants monsieur Antoine LEJAY et madame Madalina DEACONU pour leurs très grande disponibilité et leurs conseils précieux qu'ils n'ont cessé de manifester dans leurs directions de ce travail.

Ma reconnaissance va également aux membres du jury qui m'ont fait l'honneur de juger ce travail qu'ils trouvent ici le témoignage de ma profonde gratitude. Mes vifs remerciements à mes amis, qui m'ont soutenu et que je n'ai pas cité. Un grand merci à tous ceux qui ont contribué de près ou de loin à la réalisation du présent projet.

Enfin je dédie ce projet de fin d'étude à mes parents, mon frère et ma grande famille.

Table des matières

1. Présentation générale du projet	13
1.1. Contexte du projet	13
1.2. Méthodologie du travail	14
2. Mesures de risques : VaR et CVaR	17
2.1. Les outils de gestion de risque et la mesure de risque financier	17
2.2. Value at Risk (VaR)	19
2.3. Conditional Value at Risk (CVaR)	20
3. Estimation de la VaR et de la CVaR	23
3.1. L'indice boursier CAC40	23
3.2. Environnement de travail	23
3.3. Modèle d'estimation des mesures de risques	23
3.4. Estimation de la volatilité suivant le modèle EWMA	25
3.5. Calcul de la VaR paramétrique	28
3.6. Backtesting	36
3.7. Calcul de la VaR historique	38
3.8. Calcul de la CVaR paramétrique	39
3.9. Calcul de la CVaR historique	39
3.10. Inconvénient du modèle normal	40
4. Théorie des valeurs extrêmes (TVE)	43
4.1. Loi de Pareto	44
4.1.1. Estimateur de Hill	45
4.2. Approche des quantiles extrêmes par la méthode des excès	48
4.2.1. Distribution des excès des pertes	48
4.2.2. Distribution de Pareto Généralisée (GPD)	50
4.2.3. Estimation de la VaR et la CVaR	52
4.2.4. Maximum de vraisemblance (MLE)	55
4.2.5. QQ-Plot	57
A. Portail décisionnel	63
A.1. Présentation de l'application	63
A.1.1. Contexte de l'application	63
A.1.2. Démarche de réalisation de l'application BI	63
A.1.3. Objectif de l'application	64

A.2. Réalisation de l'application	64
A.2.1. Extraction des données à partir du web	64
A.2.2. Modélisation du Data Warehouse	65
A.2.3. Chargement du Data Warehouse	66
A.2.4. Réalisation des tableaux de bord	69
A.3. Conclusion	73

Table des figures

1.1. Méthodologie scrum (d'après ©Mountain Goat Software).	15
1.2. Diagramme de Gantt.	16
2.1. Représentation d'un quantile pour la loi normale.	19
3.1. Prix des actifs financiers de notre portefeuille.	25
3.2. Log-rendements des actifs financiers de notre portefeuille.	26
3.3. Variation de la volatilité du portefeuille au cours du temps estimé par la méthode EWMA.	27
3.4. La variation de la volatilité (rouge) et des log-rendements du portefeuille au cours du temps.	28
3.5. Moyenne, écart-type σ , skewness S et kurtosis K calculés sur une fenêtre mobile de deux ans pour le portefeuille introduit au paragraphe 3.3. . . .	35
3.6. L'évolution de la Value-at-Risk quotidienne à un niveau de risque de 5 % suivant le modèle normal (équ. (3.5.2)) ou avec l'approximation de Cornish-Fisher (équ. (3.5.4)).	36
3.7. Évolution du rapport entre le seuil de perte au niveau de risque $\alpha = 5\%$ et le prix calculés avec le modèle normal (noir) et l'approximation de Cornish-Fisher (rouge).	37
3.8. Histogramme des log-rendements.	40
4.1. Représentation des pertes durant les moments extrêmes.	45
4.2. Variation de l'estimateur de Hill en fonction de k avec $n = 2818$ observations.	47
4.3. Les log-pertes et le seuil $u = 2\%$, ce qui représente 165 des données (soit 5,8 % des observations).	49
4.4. La fonction de répartition des excès au delà du seuil $u = 0,02$, et comparaison avec $G_{\xi,\beta(u)}$ où $\xi = 0,97$ et $\beta = 0,007$ sont estimés par un maximum de vraisemblance.	50
4.5. Histogramme de log-pertes positives au dessus du seuil $u = 0,02$ et densité de la loi de Pareto généralisée dont les paramètres sont estimés par un maximum de vraisemblance.	51
4.6. Graphe de la moyenne des excès.	58
4.7. QQ-plot de la queue de distribution empirique contre la GPD.	59
A.1. Le processus BI.	64
A.2. Extraction des données à partir du site web.	65
A.3. Modèle du Data Warehouse.	66

A.4. Processus de l'ETL.	67
A.5. Chargement de la dimension DIM_SOCIETE.	68
A.6. Chargement de la table de fait FACT_CARNET_ORDER.	69
A.7. Exécution de tous les JOBS.	69
A.8. Processus de restitution des données.	70
A.9. Logo Tableau Software.	70
A.10. Tableau de bord du L'Oréal.	71
A.11. Tableau de bord du marché financier européen.	72
A.12. Tableau de bord de L'Oréal sur la tablette.	73
A.13. Tableau de bord du marché financier européen en ligne.	74

Liste des tableaux

3.1. Backtesting de la VaR paramétrique (loi normale).	37
3.2. Estimation de la VaR historique et paramétrique (loi normale) sur l'ensemble des données.	38
3.3. Estimation de la CVaR historique et paramétrique.	40
4.1. Estimateur de Hill pour les log-pertes.	46
4.2. Estimation des paramètres de la GPD et de la VaR quotidienne calculée sur l'ensemble des données.	53
4.3. Estimation de la $CVaR_\alpha$ pour un seuil $u = 0,02$	55

INTRODUCTION

Le monde de la finance a vécu plusieurs moments de crise dont la dernière a été déclenchée en Amérique en 2008. Depuis, le monde vit une instabilité financière. On ne peut désormais pas empêcher la situation économique d'un pays de se dégrader lorsqu'il s'agit d'un événement rare. Les événements extrêmes sont caractérisés par leur faible probabilité. Une fois apparus, les événements extrêmes procurent aux financiers des pertes importantes. Cela est dû aux modélisations de risque ordinaires qui n'évaluent pas bien les impacts de ces moments de crise. Ce qui explique que depuis quelques années, l'étude des indicateurs de risque et de la *théorie des valeurs extrêmes* est devenue un axe de recherche actif.

La gestion de risque financier est la solution immédiate vers laquelle les chercheurs, économistes et gestionnaires financiers se sont orientés depuis la dernière crise. Durant les événements extrêmes, ces derniers visent à mieux étudier le marché financier en minimisant les pertes potentielles des actifs qui constituent un portefeuille. Afin d'améliorer les indicateurs de risque actuels, ils ont axé leurs travaux sur l'étude de la queue de distribution de ces pertes.

Notre projet s'insère dans ce cadre. Notre étude portera sur l'analyse des indicateurs utilisés afin de mesurer le risque du marché d'un portefeuille d'instruments financiers. L'objectif du stage est de se familiariser avec les indicateurs utilisés dans les institutions financières pour gérer le risque. Nous commencerons par introduire les modèles de mesure de risque. Nous introduisons la *Value at Risk* (VaR) et la *Conditional Value at Risk* (CVaR). Nous calculons ces deux indicateurs par plusieurs méthodes. Par la suite, nous étudions la *Théorie des Valeurs Extrêmes* (TVE) afin de mieux décortiquer la queue de la distribution et donc avoir une vue plus claire sur les événements extrêmes. Par la suite, nous construisons un portefeuille d'actions afin de simuler numériquement ces indicateurs et étudier leurs comportements.

Dans une annexe qui correspond à un travail de développement logiciel demandé dans par mon école, nous avons implémenté un portail décisionnel qui permet de donner une vision plus claire, actualisée sur la fluctuation du marché financier.

Ce rapport s'articule autour de quatre chapitres. Le premier chapitre est consacré à la présentation générale du projet et à l'analyse du contexte de ce dernier ainsi qu'à la définition de la méthodologie du travail et la mise en place d'un plan. Le deuxième chapitre présente les mesures de risques et la définition des indicateurs financiers VaR et CVaR. Le troisième chapitre porte sur l'estimation de ces derniers selon le modèle normal. Le quatrième chapitre illustre l'étude des valeurs extrêmes des événements rares. La dernière partie en annexe illustre le portail décisionnel.

1. Présentation générale du projet

Introduction

Dans ce chapitre, nous nous proposons de décrire le contexte de notre projet. Nous présentons également l'organisme d'accueil ainsi que la méthodologie du travail.

1.1. Contexte du projet

Nous ne pouvons considérer une économie consistante sans qu'elle ait un système de maîtrise de risque fiable et rigoureux. Les pertes au niveau des portefeuilles d'actions reflètent une faille au niveau de la modélisation classique de risque. En effet, les techniques usuelles de mesure de risque ne sont pas capables de décrire réellement le marché financier aux moments extrêmes.

Suite à la dernière crise financière, les chercheurs ont accentué leurs travaux sur les mathématiques financières afin d'améliorer les techniques utilisées dans la mesure de risque. Nous nous proposons dans ce stage d'étudier une solution basée sur des indicateurs simples et significatifs afin de mieux gérer les moments de crise.

Le stage se déroule au sein de l'équipe *Tosca* d'*Inria*. Cette équipe est spécialisée dans les mathématiques appliquées, méthodes numériques probabilistes et la modélisation stochastique.

Afin de définir une mesure de risque rigoureuse des marchés financiers, nous nous proposons d'étudier l'évolution du marché financier à travers plusieurs modèles : le modèle *gaussien*, le modèle de *Pareto généralisé* de la *théorie des valeurs extrêmes*.

Notre projet portera, dans un premier temps, sur l'étude de risque d'un portefeuille : nous commençons par l'analyse mathématique des indicateurs qui le détectent. Ensuite, nous passons au calcul de ces derniers et à l'implémentation d'une solution numérique qui vise à aider les institutions financières et à simplifier leurs analyses. Nous calculons les VaR et CVaR par plusieurs méthodes : historique et paramétrique et suivant les modèles cités précédemment. Nous mesurons également l'épaisseur de la queue de la distribution des rendements financiers grâce à l'*estimateur de Hill*. La queue de la distribution est épaisse pendant les moments de crise, c'est pour cela nous privilégions nos études sur cette dernière.

Dans un second temps, nous implémentons un portail décisionnel afin de faciliter aux gestionnaires de risque la tâche de prendre décision. Cette application repose sur la *Business intelligence* (BI). Nous extrayons les données à partir d'un site internet. Nous leur appliquons le processus BI (passant de l'ETL jusqu'à la Data Warehouse). Enfin, nous montons les tableaux de bords qui illustrent l'évolution des cours boursiers.

1.2. Méthodologie du travail

Notre projet de fin d'étude comporte deux volets : le premier consacré à l'étude mathématique financière des indicateurs de risque afin de gérer le risque d'un portefeuille. Le second est consacré à la mise en place d'un portail décisionnel afin d'aider les gestionnaires des portefeuilles à suivre la variation des cours boursiers, détecter les changements brusques pour prendre les précautions nécessaires et saisir les opportunités pour générer un gain maximal. Par conséquent, il faut utiliser une méthodologie de travail qui soit adaptée aux différentes parties du projet et qui assure le bon déroulement du travail ainsi que la fluidité du processus. Cette méthodologie doit garantir :

1. La gestion de la complexité technologique.
2. L'organisation des ressources.
3. La prévention contre les dérives et les mauvaises estimations qui vont allonger les coûts et le temps de mise en place du système.
4. Le manque d'intégration entre les composantes de l'architecture du business intelligence.
5. Les limites de certains indicateurs financiers.

Face au grand nombre des méthodologies de travail (cycle en V, méthodologie Wonderfull,...), nous avons opté pour la méthodologie SCRUM qui assure :

1. La forte implication des utilisateurs tout au long du cycle du projet.
2. La capacité à livrer tôt et régulièrement (cycles itératifs et incrémentaux) des applications BI opérationnelles sur des données réelles.
3. La capacité d'introduire des changements en cours du projet pour satisfaire les besoins du client.

Pour ce faire, notre démarche de gestion de projet décisionnel en mode agile s'appuie sur cinq grands principes :

1. L'itération et l'incrémentation : expliquent le fait que notre projet soit conçu et construit par des petits morceaux successifs vus comme des points de synchronisation et de feedback entre les acteurs du projet.
2. La planification des tâches (fonctionnalités) qui sont priorisées par leurs importances métier.
3. L'auto ajustement du client peut moduler ses exigences, en connaissance de cause.
4. La mise en évidence de la communication car tous les acteurs collaborent en continu sur le projet.
5. La gestion de la conception et les plannings par toute l'équipe de façon collective à travers des réunions (Réunion de planification de sprint, réunion de révision sprint, réunion de rétrospective ...).

Du fait de l'implication de seulement 3 intervenants dans le projet, nous avons choisi d'appliquer les approches suivantes de la méthodologie scrum :

1.2. Méthodologie du travail

1. Planification du Sprint : Durant cette réunion, nous avons sélectionné les éléments prioritaires du « Product Backlog » (liste ordonnée des exigences fonctionnelles et non fonctionnelles du projet) que nous pensons pouvoir réaliser au cours du sprint.
2. Revue de Sprint : Au cours de cette réunion qui a lieu à la fin du sprint, nous avons présenté les fonctionnalités terminées au cours du sprint et recueilli les feedbacks du Product Owner et des utilisateurs finaux. C'est également le moment d'anticiper le périmètre des prochains sprints et d'ajuster au besoin la planification du release.
3. Rétrospection de Sprint : La rétrospective effectuée après la revue de sprint est l'occasion de s'améliorer (productivité, qualité, efficacité, conditions de travail ...).



FIGURE 1.1. – Méthodologie scrum (d'après ©Mountain Goat Software).

Nous détaillons dans la figure 1.2 le plan que nous avons suivi durant notre travail. Le diagramme de Gantt résume notre démarche. Étant donné que notre projet est à la base fondé sur les mathématiques financières, nous avons commencé le travail par une révision des probabilités et des processus stochastiques afin de nous familiariser avec les quantiles et mieux comprendre l'indicateur VaR. Nous avons également étudié l'historique de la VaR, son utilité et ses avantages. Afin d'avoir un indicateur rigoureux, nous avons étudié les solutions apportées pour pallier aux inconvénients de la VaR. Durant notre étude, nous avons constaté que le modèle gaussien ne capte pas les moments extrêmes. De ce fait, nous sommes passés à l'étude de la théorie des valeurs extrêmes : le modèle de *Pareto généralisé* afin d'accentuer notre travail sur la queue de la distribution qui est épaisse durant les moments de crise. Nous avons pris en considération le dépassement de la VaR et avons notamment étudié et implémenté la solution numérique de la VaR et la CVaR. Ensuite, nous avons implémenté le portail décisionnel afin de faciliter la tâche de prise de décision aux gestionnaires de risque.

Conclusion

L'étude préliminaire du sujet, la spécification du contexte et l'adoption de la méthode du travail nous ont permis de mieux nous mettre dans le cadre du projet.

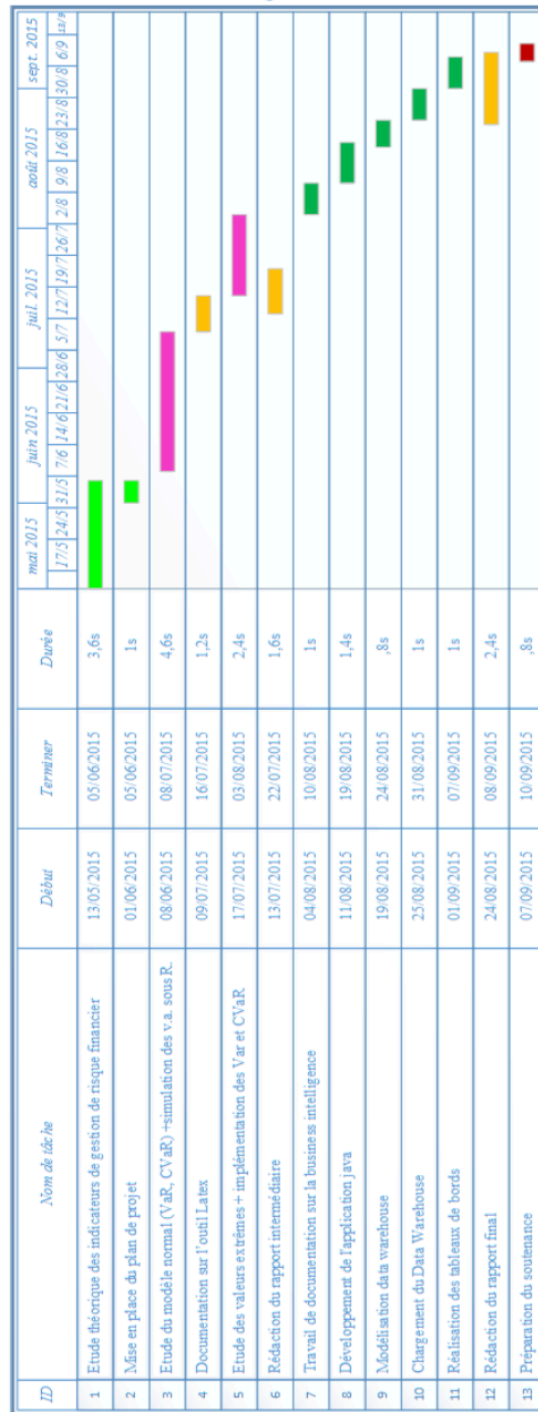


FIGURE 1.2. – Diagramme de Gantt.

2. Mesures de risques : VaR et CVaR

Introduction

La mesure de risque est devenue un enjeu majeur dans le monde de la finance. Les gestionnaires de risque et les économistes visent à résoudre la problématique de détermination du risque lié à la variation du marché financier. Nous nous intéressons dans ce chapitre à introduire deux techniques fortement recommandées aux gestionnaires de risque dans les institutions financières. Apparue pour la première fois dans les années 1990, la *Value-at-Risk* a été introduite dans la banque JP Morgan suite aux différentes crises financières que le monde a subi depuis ce temps là. Elle constitue un modèle de mesure de risque très répandu dans le monde de la finance et l'assurance. La *Value-at-Risk* est généralement accompagnée par l'*Expected Shortfall* qui est égal à la *Conditional Value at Risk* dans le cas continu.

2.1. Les outils de gestion de risque et la mesure de risque financier

L'économie mondiale se bat pour résister aux désastres dus aux moments de crise. Depuis les dernières années, la gestion de risque financier est devenue la préoccupation des managers et les institutions financières visent à maintenir leurs positions hors risque.

La mesure de risque est la perception de l'aléa associé aux instruments financiers dans lesquelles les institutions financières investissent. Nous nous référons à l'article d'Acerbi et Tasche [AT02b] afin de définir une mesure de risque cohérente.

Mesure de risque cohérente

Les accords de Bâle évoluent régulièrement afin de renforcer la sécurité et la fiabilité des systèmes financiers surtout après la dernière crise financière. Nous pouvons voir cet impact sur la gestion de risque financier qui s'est considérablement développée pointant les failles des anciennes techniques. Dans ce cadre, le développement des recherches en mathématiques financières ont prouvé qu'une mesure de risque n'est significative que si elle est cohérente.

Définition 2.1.1 (Mesure de risque). Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé représentant le risque. Considérons $\alpha \geq 0$, $a \in \mathbb{R}$, X et Y deux variables aléatoires réelles. Soit r_f le rendement financier d'un placement sans risque. Une fonction $\rho : \Omega \rightarrow \mathbb{R}$ est dite *une mesure de risque cohérente* si elle vérifie les quatre axiomes suivants :

1. La monotonie : si $X \geq Y$, alors $\rho(X) \leq \rho(Y)$.

2. L'homogénéité : $\rho(aX) = a\rho(X)$.
3. L'invariance par translation : $\rho(X + ar_f) = \rho(X) - a$.
4. La sous-additivité : $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

Interprétation des axiomes

1. La monotonie traduit le fait que si le risque associé à un actif est plus grand qu'un autre, alors le capital nécessaire à couvrir ce dernier est inférieur à l'autre.
2. L'homogénéité traduit le fait que le risque associé à un actif est proportionnel à son poids.
3. L'invariance par translation traduit le fait que si on ajoute un montant a au risque associé au portefeuille, sa mesure de risque diminue du même montant a . Autrement dit, $\rho(X + \rho(X)r_f) = \rho(X) - \rho(X) = 0$. On en déduit donc le fait que la mesure de risque représente le capital (la quantité d'investissement sans risque) que l'on doit ajouter à son portefeuille afin de le rendre globalement sans risque.
4. La sous-additivité traduit la diversification dans un portefeuille : la mesure de risque d'un ensemble d'actifs est inférieure à la somme des mesures de risque de ces derniers séparément.

Quantiles

Nous allons très souvent utiliser la notion du quantile dans la suite parce que les quantiles sont la base de la modélisation financière du risque. Ils représentent le rendement potentiel des actifs à une probabilité donnée sur un intervalle de temps déterminé.

Définition 2.1.2. Soit $\alpha \in [0, 1]$ et X une variable aléatoire continue. On appelle *quantile* de X , la valeur que prend la fonction de répartition inverse associée à cette dernière pour une probabilité α donnée. On note ce quantile q_α tel que :

$$\begin{aligned} q_\alpha(X) &= \inf\{x \in \mathbb{R} ; \mathbb{P}[X \leq x] \geq \alpha\} \\ &= F_X^{-1}(\alpha), \end{aligned}$$

où F_X^{-1} est la fonction de répartition inverse associée à X si elle existe.

Pour α proche de 1, le quantile $q_\alpha(X)$ représente le gain maximal (par rapport au prix ou au rendement) que peut avoir un investisseur sur un horizon bien déterminé et à un seuil α préalablement fixé. Les économistes, gestionnaires de risque et les financiers s'intéressent généralement à l'opposé de ce quantile afin de prédire la perte maximale qu'ils peuvent subir. Ils visent ainsi à déterminer les dépassements potentiels par rapport au seuil α que leurs portefeuilles sont susceptibles de subir sur un intervalle de temps donné.

2.2. Value at Risk (VaR)

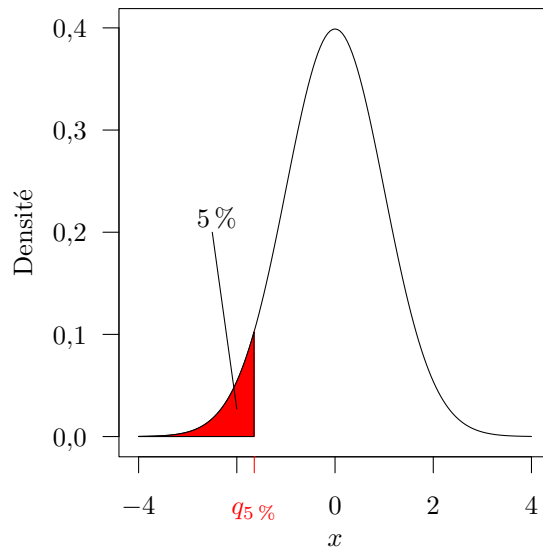


FIGURE 2.1. – Représentation d’un quantile pour la loi normale.

2.2. Value at Risk (VaR)

La *Value at Risk* ou plus simplement VaR, est un outil de gestion de risque utilisé dans les institutions financières. C’est une mesure de risque représentant la perte potentielle maximale que peut subir un portefeuille à un niveau de probabilité préalablement fixé sur une période de temps bien déterminée. Cet outil correspond aux recommandations des organismes régulateurs. Il dépend de la loi de distribution des rendements financiers. Nous introduisons la VaR comme elle a été définie dans l’article de Jaroslav et Baran [BW12].

Définition 2.2.1. Soit X une variable aléatoire représentant le rendement d’un instrument financier. Elle est considérée en tant que gain si elle est positive et perte si elle est négative. Soit $\alpha \in [0, 1]$ le seuil par rapport auquel on souhaite déterminer la perte potentielle (Si α est proche de 1 c’est un seuil de confiance sinon il s’agit d’un niveau de risque). La *Value-at-Risk* est donnée par :

$$\text{VaR}_{\alpha}^{+}(X) = -\inf\{x \in \mathbb{R}; \mathbb{P}(X \leq x) > \alpha\} = q_{1-\alpha}(-X).$$

Ici, VaR^{+} est par définition une quantité positive. Il est parfois naturel et pratique d’utiliser $\text{VaR} = -\text{VaR}^{+}$, qui sera donc négative et qui peut se comparer aux pertes (signées).

La figure 2.1 donne l’allure de la répartition des rendements financiers constituant un portefeuille. Nous retrouvons ainsi la représentation graphique de la VaR sur un jour à une probabilité 95 % d’une variable aléatoire X qui suit la loi normale standard $N(0, 1)$. Nous rappelons que 95 % est le seuil de confiance par rapport auquel nous souhaitons déterminer les pertes potentielles de notre portefeuille d’instruments financiers. Le trait rouge coupe la ligne des abscisses en $-q_{95\%} = q_{1-95\%}$ qui représente la $\text{VaR}_{95\%}$. En valeur

absolue, il s'agit de la somme d'argent maximale que notre portefeuille est susceptible de perdre à un niveau de confiance de 95 % des cas sur une période de temps donnée (exemple : un jour).

Inconvénients de la VaR

L'étude des événements extrêmes a montré que le calcul de la VaR à l'aide du modèle gaussien ne nous permet pas de détecter les événements extrêmes. Bien que la VaR soit d'une grande aide pour les gestionnaires de risque financier, elle a plusieurs inconvénients. En particulier, ce n'est pas une mesure de risque cohérente car elle n'est pas sous-additive. Elle ne favorise pas nécessairement la diversification d'un portefeuille qui est une dimension essentielle dans la gestion de risque mise en évidence par Markowitz. Pour cela, nous allons nous intéresser à estimer la moyenne de dépassement au delà de la VaR^+ à l'aide de l'indicateur CVaR.

2.3. Conditional Value at Risk (CVaR)

Nous avons introduit dans la partie précédente la VaR. Il s'agit d'une mesure de risque très utile et assez répandue dans les institutions financières mais qui ne favorise pas la diversification du portefeuille. Donc nous ne pouvons plus nous y appuyer.

Afin de déterminer les dépassements des pertes (les rendements financiers) au delà de la VaR^+ , nous présentons la *Conditional Value at Risk* (CVaR). C'est une mesure de risque qui a été introduite comme complément à la VaR. Elle favorise la diversification des instruments financiers d'un portefeuille et traite la queue. Nous nous appuyons dans la suite sur l'article d'Acerbi et Tasche [AT02a] pour définir la CVaR, donner ses propriétés, et montrer qu'elle est sous-additive.

Définition 2.3.1 (Conditional Value at Risk et Expected Shortfall). Soit X une variable aléatoire intégrable représentant le risque sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\alpha \in [0, 1]$. Nous définissons l'*Expected Shortfall* ES_α au niveau de risque α (ou *Conditional Value at Risk* dans le cas continu) par :

$$ES_\alpha(X) = CVaR_\alpha(X) = -\mathbb{E}[X|X \leq q_\alpha(X)].$$

En d'autres termes, la CVaR représente la moyenne des pertes (positives) qui dépassent VaR^+ . Vu que la VaR^+ mesure la valeur qui sépare les $(1 - \alpha)\%$ de la distribution, nous accentuons notre recherche sur la queue de la distribution des pertes, c'est-à-dire les $\alpha\%$ restant.

Proposition 2.3.2. Soit X une variable aléatoire intégrable sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et $\alpha \in [0, 1]$ est un niveau de risque fixé. Si X représente les (log-)rendements,

$$CVaR_\alpha(X) = \mathbb{E}[X|X > VaR_\alpha(X)].$$

2.3. Conditional Value at Risk (CVaR)

Démonstration. En effet, nous avons :

$$\begin{aligned}\text{CVaR}_\alpha(X) &= -\mathbb{E}[X|X \leq q_\alpha(X)] = -\mathbb{E}[X| -X > -q_\alpha(X)] \\ &= \mathbb{E}[-X| -X > \text{VaR}_\alpha^+(X)] \text{ car } \text{VaR}_\alpha^+(X) = -q_\alpha(X).\end{aligned}$$

Ce qui conduit au résultat de la proposition 2.3.2. \square

Proposition 2.3.3. *Soit X une variable aléatoire intégrable sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et $\alpha \in [0, 1]$ fixé. Nous avons :*

$$\text{CVaR}_\alpha(X) = -\frac{\mathbb{E}[X1_{\{X \leq q_\alpha(X)\}}]}{\mathbb{P}[X \leq q_\alpha(X)]}.$$

Par la continuité sur $[0, 1]$ de la fonction $\alpha \mapsto \text{CVaR}_\alpha$, nous obtenons $\mathbb{P}[X \leq q_\alpha(X)] = \alpha$. Ainsi, nous avons :

$$\text{CVaR}_\alpha(X) = -\frac{1}{\alpha}\mathbb{E}[X1_{\{X \leq q_\alpha(X)\}}].$$

Par la suite, nous nous proposons de démontrer que la CVaR est sous additive. C'est-à-dire qu'elle favorise la diversification des actifs dans un portefeuille. Pour démontrer ceci, nous aurons besoin des résultats suivants.

Définition 2.3.4. Soient X une variable aléatoire et $\alpha \in [0, 1]$. Nous notons par $1_{\{X \leq x\}}^{(\alpha)}$ la fonction définie sur $[0, 1]$ par :

$$1_{\{X \leq x\}}^{(\alpha)} = \begin{cases} 1_{\{X \leq x\}} + \frac{\alpha - \mathbb{P}[X \leq x]}{\mathbb{P}[X = x]} 1_{\{X = x\}} & \text{si } \mathbb{P}[X = x] > 0. \\ 1_{\{X \leq x\}} & \text{si } \mathbb{P}[X = x] = 0. \end{cases}$$

Propriétés 2.3.5. *Soient X une variable aléatoire telle que $\mathbb{E}[X^-] < \infty$ sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et $\alpha \in [0, 1]$. Alors*

1. $1_{\{X \leq x\}}^{(\alpha)} \in [0, 1]$.
2. $\mathbb{E}[1_{\{X \leq x\}}^{(\alpha)}] = \alpha$.
3. $\text{CVaR}_\alpha(X) = -\frac{1}{\alpha}\mathbb{E}[X1_{\{X \leq x\}}^{(\alpha)}]$.

Contrairement à la VaR, la CVaR est bien une mesure de risque sous-additive.

Proposition 2.3.6. *Soient X et Y deux variables aléatoires intégrables ayant respectivement comme quantiles x_α et y_α . Alors*

$$\text{CVaR}_\alpha(X + Y) \leq \text{CVaR}_\alpha(X) + \text{CVaR}_\alpha(Y).$$

Démonstration. Soit $W = \alpha(\text{CVaR}_\alpha(X) + \text{CVaR}_\alpha(Y) - \text{CVaR}_\alpha(Z))$ où Z est une variable aléatoire telle que $Z = X + Y$ de quantile z_α .

$$\begin{aligned} W &= \mathbb{E}[Z1_{\{Z \leq z_\alpha\}}^{(\alpha)} - X1_{\{X \leq x_\alpha\}}^{(\alpha)} - Y1_{\{Y \leq y_\alpha\}}^{(\alpha)}] \\ &= \mathbb{E}[X(1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{X \leq x_\alpha\}}^{(\alpha)}) + Y(1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{Y \leq y_\alpha\}}^{(\alpha)})] \\ &= \mathbb{E}[X(1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{X \leq x_\alpha\}}^{(\alpha)})] + \mathbb{E}[Y(1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{Y \leq y_\alpha\}}^{(\alpha)})] \\ &= W_1 + W_2, \end{aligned}$$

avec $W_1 = \mathbb{E}[X(1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{X \leq x_\alpha\}}^{(\alpha)})]$ et $W_2 = \mathbb{E}[Y(1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{Y \leq y_\alpha\}}^{(\alpha)})]$. Or d'après la définition 2.3.4, si nous remplaçons x par x_α nous avons :

1. Si $X > x_\alpha \Rightarrow 1_{\{X \leq x_\alpha\}}^{(\alpha)} = 0$ alors $1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{X \leq x_\alpha\}}^{(\alpha)} \geq 0$.
2. Si $X \leq x_\alpha \Rightarrow 1_{\{X \leq x_\alpha\}}^{(\alpha)} = 1$ alors $1_{\{Z \leq z_\alpha\}}^{(\alpha)} - 1_{\{X \leq x_\alpha\}}^{(\alpha)} \leq 0$.

Donc $W_1 \geq 0$. Avec une manipulation similaire sur la variable aléatoire Y , nous démontrons que $W_2 \geq 0$. Ainsi $W = W_1 + W_2 \geq 0$. Cela implique que $\text{CVaR}_\alpha(X) + \text{CVaR}_\alpha(Y) \geq \text{CVaR}_\alpha(Z) \geq \text{CVaR}_\alpha(X + Y)$. Nous avons ainsi la sous-additivité de la CVaR. Ceci met en évidence le fait que CVaR favorise la diversification des actifs dans un portefeuille. \square

Conclusion

Nombreux sont les outils de gestion de risques financiers qui ont été mis en place au fil du temps. En effet, les recherches en mathématiques financières et les progrès effectués dans le domaine de gestion de risque ont permis de faire la migration de mesure de risque du calcul de la volatilité au calcul des indicateurs Value at Risk et la Conditional Value at Risk qui reposent sur la notion du quantile. Ces deux techniques simples permettent de suivre et de surveiller les risques conformes aux réglementations des organismes régulateurs. Dans ce chapitre nous avons introduit ces deux techniques et présenté quelques propriétés qui leurs sont liées. Nous procédons dans la suite à leurs estimations suivants plusieurs méthodes (paramétrique et historique).

3. Estimation de la VaR et de la CVaR

Introduction

Dans ce chapitre nous allons calculer la VaR et CVaR par plusieurs méthodes (historique et paramétriques) suivant le modèle gaussien. Nous allons également introduire l'indice boursier CAC40 sur lequel nous allons mener notre simulation. Nous commençons d'abord par introduire quelques notions importantes des mathématiques financières afin d'aboutir aux résultats souhaités. Pour plus de détails, le lecteur peut se référer à l'article de Jaroslav et Baran [BW12].

3.1. L'indice boursier CAC40

Le CAC40 est un produit boursier conçu à partir du cours des quarante sociétés les plus actives du marché financier français. Ces sociétés ont été soigneusement choisies parmi les 100 sociétés ayant les plus grands fonds. Les cotations des cours de ces actions se fait d'une manière continue : c'est-à-dire que la mise à jours de la valeur de ces cours se fait toutes les 30 secondes. Ainsi, cet indicateur reflète la tendance globale de l'économie française.

3.2. Environnement de travail

Nous utiliserons le logiciel **R** comme environnement de travail afin de simuler le risque, prédire les pertes potentielles et analyser les graphes de fluctuation des prix sur le marché financier. **R** est un logiciel qui offre des solutions statistiques. Sa puissance dans le domaine de l'analyse des données, repose sur la multitude des packages développés et mis librement à la disposition de son utilisateur. Ces paquets couvrent des nombreuses fonctionnalités dans des divers domaines tels que l'économétrie, gestion de risque, série temporelle. Grâce à cette diversité de fonctionnalités, **R** est considéré comme la référence de l'analyse des données et des statistiques. Un des atouts de **R** est la facilité de production des graphiques de bonne qualité. Nous pouvons ainsi manipuler aisément les équations mathématiques compliquées.

3.3. Modèle d'estimation des mesures de risques

En finance du marché, il est essentiel d'estimer les mesures de risques afin de modéliser les rendements futurs. Pour ce faire, nous nous basons sur les volatilités et les rendements passés tout en supposant que ces derniers sont normalement distribués.

Définition 3.3.1 (Rendement logarithmique d'un instrument financier). Soit P_t le prix de l'actif à la date t (ici, une unité représente un jour). On appelle *rendement logarithmique continu journalier* des actifs la suite $(X_t)_{t \in \mathbb{N}}$ tel que $X_t = \log(\frac{P_t}{P_{t-1}})$. Nous supposons que les rendements $(X_t)_{t \in \mathbb{N}}$ sont normalement distribués conditionnellement aux informations de la date t tels que : $X_t = \sigma_t \varepsilon_t \sim N(0, \sigma_t^2)$ où $(\varepsilon_t)_{t \in \mathbb{N}}$ est une suite de variable aléatoire i.i.d. telle que $\mathbb{E}[\varepsilon_t] = 0$ et $\text{Var}(\varepsilon_t) = 1$.

Les rendements financiers sont la base de l'étude de la gestion de risque. C'est à partir de ces derniers que nous déterminons les variations du marché financier. Ils décrivent l'évolution de l'instrument financier au cours du temps.

Définition 3.3.2 (Rendement logarithmique de plusieurs instruments financiers). On appelle *rendement d'un portefeuille* constitué de n instruments financiers à la date t (t exprimé en unité de jour) la somme X_t^p tel que $X_t^p = \sum_{i=1}^n \omega_i X_{i,t}$ où ω_i est le poids correspondant à l'actif i .

Dans notre stage, nous avons choisi de travailler sur un portefeuille constitué de deux actifs. Nous avons travaillé sur le prix des actions Air Liquide et Sanofi du CAC40, sur la période allant du 1/1/2004 au 31/12/2014, soit 10 ans. La stratégie de composition de ce dernier est d'investir ω_1 actions du premier actif et ω_2 actions du second. Donc, notre portefeuille à l'instant t est comme suit :

$$V_t = \omega_1 P_{1,t} + \omega_2 P_{2,t}$$

Lorsque nous nous proposons d'étudier un portefeuille constitué de plusieurs instruments financiers, nous devons considérer le rendement de ces derniers afin de déterminer sa variation au cours du temps. La fluctuation des rendements décrit des pertes ou bien des gains. Les figures 3.1 et 3.2 représentent les prix les rendements logarithmiques journaliers des deux actifs du portefeuille que nous avons construit (constitué de deux actifs). Les pics que nous remarquons sur les graphes de (3.2) reflètent des moments de crise.

Dans les simulations numériques, nous avons utilisé la composition $\omega_1 = 2$ et $\omega_2 = 1$.

Nous essayons dans la suite de représenter les rendements logarithmiques avec la volatilité afin de voir la correspondance entre les pics de chacune au cours du temps.

Pour certains financiers qui cherchent un gain sur une courte durée de temps, on parle généralement des spéculateurs, la volatilité constitue un critère d'investissement. On note par σ_t la volatilité d'un actif. Elle dépend du temps.

Définition 3.3.3 (Variance d'un portefeuille). On appelle *variance d'un portefeuille* à la date t la quantité $\sigma_{p,t}^2$ telle que $\sigma_{p,t}^2 = \omega^t \Sigma_t \omega$, avec ω le vecteur des poids des actifs constituant le portefeuille et Σ_t la matrice de covariance du portefeuille entre deux actifs i et j à la date t telle que $\Sigma_t = (\sigma_{ij,t}^2)_{1 \leq i,j \leq n}$.

Proposition 3.3.4. Si les rendements sont normalement distribués avec $X_t \sim N(0, \sigma_t^2)$ alors la combinaison linéaire de ces derniers est aussi normalement distribuée, conditionnellement à la date t , et $X_t^p \sim N(0, \sigma_{p,t}^2)$ où $\sigma_{p,t}^2 = \omega^t \Sigma_t \omega$.

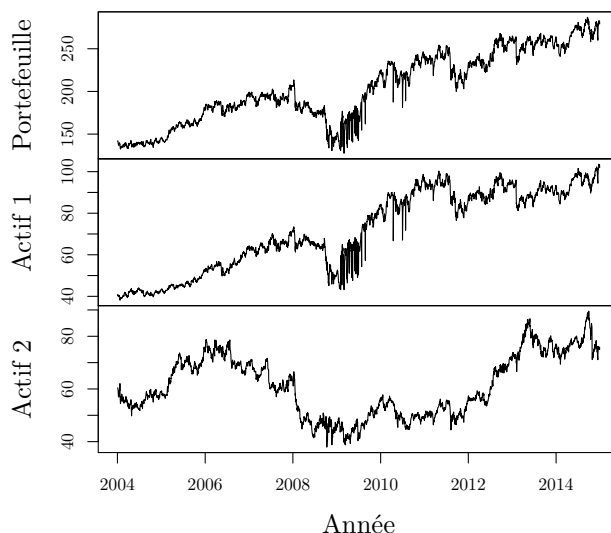


FIGURE 3.1. – Prix des actifs financiers de notre portefeuille.

Avant l'apparition des indicateurs VaR et CVaR, les gestionnaires et les économistes ont toujours utilisé la variance et l'écart type comme étant des mesures de risques. La variation de ces derniers au cours du temps décrit la fluctuation du marché financier. Nous avons choisi la méthode du lissage exponentiel pour étudier la variance du marché financier au cours du temps.

3.4. Estimation de la volatilité suivant le modèle EWMA

Afin de déterminer la variance du rendement d'un actif ainsi que sa covariance, nous nous basons sur les modèles auto-régressifs. Ces modèles se basent sur le fait que les données passées jusqu'à la date t peuvent décrire le comportement à la date $t + 1$. Nous considérons un échantillon de données passées de taille n . Nous supposons que les rendements sur lesquels nous allons travailler sont normalement distribués ($X_t \sim N(0, \sigma_t^2)$) et nous appliquons le modèle *Exponentially Weighted Moving Average* (EWMA), utilisé dans le RiskMetrics [Lon94] et qui sert à prédire la volatilité et la variance d'un portefeuille à partir des données passées. Ce modèle vise à donner un grand poids aux données récentes. Cette attribution de poids sera accordée aux rendements financiers d'une manière proportionnelle. On affecte au rendement le plus récent un poids égal à $(1 - \lambda)$. Le jour juste avant on lui attribue $\lambda(1 - \lambda)$ et ainsi de suite. Le λ est le facteur de lissage. Nous nous appuyons sur ce modèle pour estimer la volatilité et la covariance futures des rendements logarithmiques journaliers. Nous obtenons donc les résultats suivants :

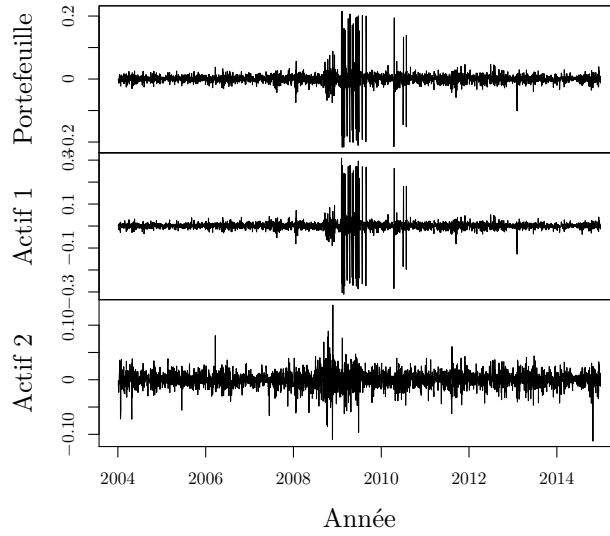


FIGURE 3.2. – Log-rendements des actifs financiers de notre portefeuille.

Proposition 3.4.1. *La variance du rendement d'un actif j est donnée par :*

$$\sigma_{j,t+1|t}^2 = \mathbb{E}_t[X_{j,t+1}^2] = (1 - \lambda)X_{j,t}^2 + \lambda\sigma_{j,t}^2, \quad \forall j = 1, \dots, n,$$

où \mathbb{E}_t désigne l'espérance conditionnelle basée sur les informations jusqu'à la date t et $\sigma_{j,t+1|t}^2$ désigne la prévision de la variance du jour $t + 1$ suivant les informations passées jusqu'à la date t . λ est le facteur de lissage, qui est une constante entre 0 et 1.

Démonstration. Nous estimons la variance en $t + 1$ d'un actif j à l'aide des rendements passés jusqu'à la date t .

$$\begin{aligned} \sigma_{j,t+1|t}^2 &= (1 - \lambda) \sum_{k=1}^N \lambda^k X_{j,t-k+1}^2 \\ &= (1 - \lambda)(\lambda^0 X_{j,t}^2 + \lambda^1 X_{j,t-1}^2 + \lambda^2 X_{j,t-2}^2 + \dots + \lambda^N X_{j,t-N+1}^2) \\ &= \dots \\ &= (1 - \lambda)X_{j,t}^2 + \lambda(1 - \lambda)(X_{j,t-1}^2 + \dots + \lambda^{N-1} X_{j,t-N+1}^2). \end{aligned}$$

D'où le résultat. □

Proposition 3.4.2. *La covariance du rendement entre deux actifs i et j est donnée par :*

$$\sigma_{ij,t+1|t}^2 = \mathbb{E}_t[X_{i,t+1}X_{j,t+1}] = (1 - \lambda)X_{j,t}X_{i,t} + \lambda\sigma_{ij,t}^2, \quad \forall i, j = 1, \dots, n.$$

3.4. Estimation de la volatilité suivant le modèle EWMA

Démonstration. Nous estimons la covariance en $t + 1$ des actifs i et j à l'aide des rendements passés jusqu'à la date t :

$$\begin{aligned}\sigma_{ij,t+1|t}^2 &= (1 - \lambda) \sum_{k=1}^N \lambda^k X_{i,t-k+1} X_{j,t-k+1} \\ &= (1 - \lambda)(X_{i,t} X_{j,t} + \lambda X_{i,t-1} X_{j,t-1} + \dots + \lambda^N X_{i,t-N+1} X_{j,t-N+1}) \\ &= \dots \\ &= (1 - \lambda) X_{i,t} X_{j,t} + \lambda(1 - \lambda)(X_{i,t-1} X_{j,t-1} + \dots + \lambda^{N-1} X_{i,t-N+1} X_{j,t-N+1}).\end{aligned}$$

D'où le résultat. \square

Remarque 3.4.3. Afin de fournir la bonne qualité de prévision, nous devons être rigoureux dans le choix de λ . Ce coefficient de lissage est très important. Il conditionne la prévision future à travers le degré de pondération que l'on affecte aux carrés des rendements. Dans la pratique, nous choisissons le $\lambda = 0,94$ comme proposé dans RiskMetrics.

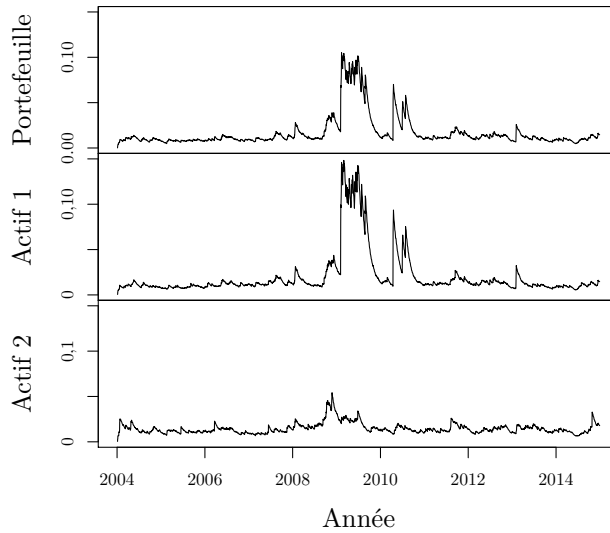


FIGURE 3.3. – Variation de la volatilité du portefeuille au cours du temps estimée par la méthode EWMA.

Les courbes de la figure 3.3 représentent la variation de la volatilité au cours du temps de notre portefeuille constitué de deux cours du CAC 40 suivant le modèle de EWMA.

Nous pouvons remarquer sur la figure 3.4 la synchronisation de la variation des log-rendements et l'évolution de la volatilité au cours du temps. En effet, chaque pic de la volatilité correspond à un pic au niveau du rendement. Nous en déduisons que la volatilité

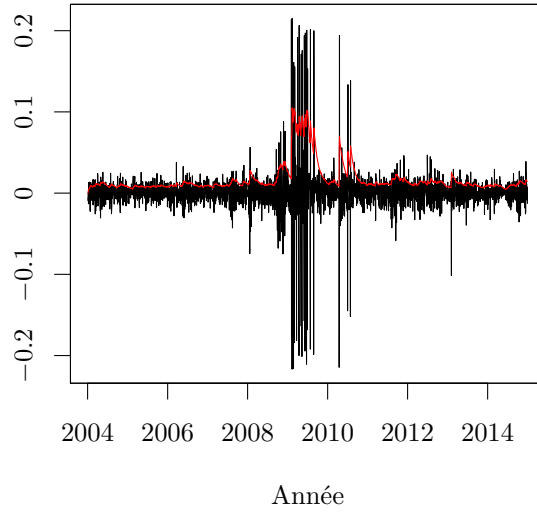


FIGURE 3.4. – La variation de la volatilité (rouge) et des log-rendements du portefeuille au cours du temps.

n'est pas une mesure de risque fiable car elle nous indique seulement l'ampleur des pics et pas leur sens (nous ne pouvons pas savoir s'il s'agit bien d'un gain ou bien d'une perte).

Définition 3.4.4. Nous supposons que notre portefeuille est composé de deux actifs i et j . Si ces deux derniers évoluent dans le même sens, autrement dit leurs rendements augmentent ou diminuent au même temps, alors nous additionnons leurs pertes/gains. Si ils varient dans deux sens opposés, alors la perte sur l'un sera compensée par le gain sur l'autre et donc notre mesure sur le risque sera nulle. Cette variation est traduite par le coefficient de corrélation entre les deux actifs $\rho_{ij,t+1|t} = \frac{\sigma_{ij,t+1|t}^2}{\sigma_{i,t+1|t}\sigma_{j,t+1|t}}$ où $\sigma_{j,t+1|t} = \sqrt{\sigma_{j,t+1|t}^2}$ est la volatilité du rendement $X_{j,t+1}$.

3.5. Calcul de la VaR paramétrique

Instrument financier

Afin de pouvoir calculer la VaR paramétrique, nous devons distinguer la nature de l'instrument que nous allons traiter et qui peut être soit un instrument financier linéaire (actions) ou bien non-linéaire (options). Après avoir su la nature de l'actif constituant le portefeuille, nous appliquons la série de Taylor suivant Pichler et Selitsch [PS99].

3.5. Calcul de la VaR paramétrique

Définition 3.5.1. Soient I un intervalle sur \mathbb{R} , a un point intérieur à I et $f : I \rightarrow \mathbb{R}$ une fonction. On dit que f est une série de Taylor au point a si et seulement si f est C^∞ . On a alors pour tout $k \geq 1$

$$f(x_1, x_2, \dots, x_k) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} \frac{\partial^{n_1}}{\partial x_1^{n_1}} \cdots \frac{\partial^{n_k}}{\partial x_k^{n_k}} \frac{f(a_1, \dots, a_k)}{n_1! \cdots n_k!} (x_1 - a_1)^{n_1} \cdots (x_k - a_k)^{n_k}. \quad (3.5.1)$$

S'il s'agit d'un instrument financier linéaire (actions) nous appliquons le développement de Taylor au premier ordre. Si c'est un instrument financier non-linéaire (options) nous appliquons le développement de Taylor au second ordre.

Calcul de VaR linéaire — l'approche par calcul de delta

Soit $i \in \{1, \dots, n\}$. Nous désignons par X_i le rendement logarithmique de cet instrument, $V(P_1, P_2, \dots, P_n)$ un portefeuille composé de n prix d'instruments financiers linéaires $(P_i)_{1 \leq i \leq n}$. Le rendement d'un portefeuille est un changement du prix initial de ce dernier au cours du temps. Ce changement est souvent utilisé comme outil de mesure de risque.

$$\begin{aligned} X_{i,t} &= \log \left(\frac{P_{i,t}}{P_{i,t-\Delta t}} \right) = \log \left(\frac{P_{i,t} - P_{i,t-\Delta t} + P_{i,t-\Delta t}}{P_{i,t-\Delta t}} \right) = \log \left(\frac{\Delta P_{i,t}}{P_{i,t-\Delta t}} + 1 \right) \\ &\simeq \frac{\Delta P_{i,t}}{P_{i,t-\Delta t}} \text{ au voisinage de zéro.} \end{aligned}$$

Instrument financier linéaire

Un instrument financier est dit linéaire si le changement de la valeur de son rendement ΔP_i sur chaque période $(t, t + \Delta t)$ est linéaire. Donc le développement de Taylor au premier ordre au portefeuille V aboutit à :

$$\begin{aligned} \frac{V(P_{1,t}, \dots, P_{i,t} + h, \dots, P_{n,t}) - V(P_{1,t}, \dots, P_{n,t})}{h} &= \frac{\partial V}{\partial P_{i,t}} \\ \text{et donc } V(P_{1,t}, \dots, P_{i,t} + h, \dots, P_{n,t}) - V(P_{1,t}, \dots, P_{n,t}) &= h \frac{\partial V}{\partial P_{i,t}}. \end{aligned}$$

Nous écrivons la légère variation h du prix de l'actif i sous la forme $h = \rho P_{i,t}$. On a alors :

$$V(P_{1,t}, P_{2,t}, \dots, (1 + \rho)P_{i,t}, \dots, P_{n,t}) - V(P_{1,t}, P_{2,t}, \dots, P_{n,t}) = \rho P_{i,t} \frac{\partial V}{\partial P_{i,t}}.$$

Nous exprimons le coefficient ρ en fonction du rendement à l'instant $t + 1$. Sachant que le nouveau prix à l'instant $t + 1$ après la variation est :

$$P_{i,t+1} \simeq (1 + \rho)P_{i,t} \Rightarrow \rho = \log \left(\frac{P_{i,t+1}}{P_{i,t}} \right) \simeq X_{i,t+1}.$$

Remarque 3.5.2. Afin d'alléger les notations, nous notons dans la suite la variation du portefeuille au cours du temps par ΔV qui dépend du temps t au lieu de $\Delta V_t = V_{t+1} - V_t$.

Nous pouvons ainsi écrire la variation du portefeuille au cours du temps en tant que la somme suivante :

$$\Delta V = \sum_{i=1}^n \frac{\partial V}{\partial P_{i,t}} \Delta P_{i,t+1} = \sum_{i=1}^n \frac{\partial V}{\partial P_{i,t}} P_{i,t} \frac{P_{i,t+1} - P_{i,t}}{P_{i,t}} \simeq \sum_{i=1}^n \delta_{i,t} X_{i,t+1},$$

où $\delta_{i,t} = \frac{\partial V}{\partial P_{i,t}} P_{i,t}$ est appelée *delta du rendement ajusté*. C'est la sensibilité du portefeuille par rapport aux actions qui le constituent. Cette sensibilité est une variation de la valeur de l'actif lorsqu'une autre donnée varie en même temps d'une quantité h très petite. Ainsi en pratique, nous rapprochons la dérivée $\frac{\partial V}{\partial P_{i,t}}$ par un changement au niveau du prix $P_{i,t}$. Nous considérons une variation $Y_i = \exp(X_i)$. Nous avons dans ce cas :

$$\begin{aligned} \Delta V &= V(P_1 Y_1, P_2 Y_2, \dots, P_{i,t} Y_i, \dots, P_n Y_n) - V(P_1, P_2, \dots, P_{i,t}, \dots, P_n) \\ &= \sum_{i=1}^n \frac{\partial V(P_{i,t} Y_i - P_{i,t})}{\partial P_{i,t}} \\ &= \sum_{i=1}^n \frac{\partial V}{\partial P_{i,t}} P_{i,t} (Y_i - 1). \end{aligned}$$

Comme les log-rendements $(X_i)_{1 \leq i \leq n}$ sont proches de zéro, nous pouvons faire l'approximation suivante au voisinage de zéro :

$$Y_i - 1 = \exp(X_i) - 1 \simeq X_i + 1 - 1 \simeq X_i.$$

Donc la variation du portefeuille se réduit à :

$$\Delta V \simeq \sum_{i=1}^n \frac{\partial V}{\partial P_{i,t}} P_{i,t} X_{i,t+1},$$

puisque la variation ordinaire d'un portefeuille constitué de n instruments financiers linéaires ayant une stratégie de composition $(n_i)_{1 \leq i \leq n}$ est comme suit :

$$\Delta V = \sum_{i=1}^n n_i P_{i,t} X_{i,t+1}.$$

Alors, par linéarité, nous pouvons considérer la $\frac{\partial V}{\partial P_{i,t}}$ égale au nombre d'actions n_i de l'instrument financier linéaire i et donc la sensibilité du portefeuille n'est autre que le stratégie du portefeuille multipliée par le prix de l'instrument financier à chaque date t ($\delta_{t,i} = n_i P_{i,t}$). Nous avons obtenu cette conclusion grâce à la linéarité des instruments financiers qui constituent le portefeuille.

Calcul de VaR linéaire

Afin de pouvoir calculer la VaR linéaire il faut passer à l'écriture vectorielle tout en supposant que le portefeuille contient des instruments financiers linéaires. Soit \mathbf{X} un vecteur de n rendements financiers logarithmiques journaliers $(X_i)_{1 \leq i \leq n}$. La matrice de covariance de ces rendements est la matrice carrée Σ de dimension n . Nous considérons δ^T le vecteur de n sensibilités $\delta_{i,t}$ ($i = 1, \dots, n$). Comme nous avons mentionné précédemment, ces sensibilités sont le produit du prix de l'instrument financier linéaire i par le nombre d'action qui lui soit associé. Nous supposons que \mathbf{X} suit la loi normale $N(\mu, \Sigma)$. La variation de notre portefeuille entre la date t et la date $t+1$ suit aussi de loi gaussienne avec les paramètres suivants :

$$\Delta V = \sum_{i=1}^n n_i P_{i,t} X_i \sim N(\mu_v, \sigma_v^2),$$

où la moyenne de la variation du portefeuille est :

$$\mu_v = \sum_{i=1}^n \delta_i \mu_i = \sum_{i=1}^n n_i \mu_i P_{i,t},$$

où $(\mu_i)_{1 \leq i \leq n}$ est le vecteur moyenne des log-rendements de instrument financier i . La variance est donnée par :

$$\sigma_v^2 = \delta^T \Sigma \delta.$$

Si nous fixons notre seuil de risque à α (c'est-à-dire α proche de 0), alors la probabilité que les pertes : les rendements négatifs (entre deux jours successifs) ne dépassent pas la perte maximale potentielle sur une période de temps fixée (VaR_α) est égale à α . C'est-à-dire :

$$\mathbb{P}[\Delta V \leq \text{VaR}_\alpha] = \alpha.$$

Comme $\Delta V \sim N(\mu_v, \sigma_v^2)$ implique que $(\Delta V - \mu_v)/\sigma_v \sim N(0, 1)$, il s'ensuit que

$$\mathbb{P}\left[\frac{\Delta V - \mu_v}{\sigma_v} \leq \frac{\text{VaR}_\alpha - \mu_v}{\sigma_v}\right] = \alpha \text{ soit } \frac{\text{VaR}_\alpha - \mu_v}{\sigma_v} = z_\alpha,$$

où $z_\alpha < 0$ est le α -quantile de la distribution normale centrée réduite. Donc pour $\alpha \in [0, 1]$ la VaR au seuil α à l'instant $t+1$ est donnée par :

$$\text{VaR}_{\alpha,t+1} = \mu_v + z_\alpha \sigma_v, \quad \alpha \text{ seuil de risque.} \quad (3.5.2)$$

Nous avons utilisé le quantile de la loi normale parce que ΔV est normalement distribué.

Si nous travaillons avec α seuil de confiance (c'est-à-dire α proche de 1), alors la probabilité que les pertes ne dépassent pas la perte maximale potentielle sur une période de temps fixée (VaR_α) est égale à α . C'est à dire : $\mathbb{P}[\Delta V \leq \text{VaR}_\alpha] = \alpha$. Comme $\Delta V \sim N(\mu_v, \sigma_v)$ implique que $(\Delta V - \mu_v)/\sigma_v \sim N(0, 1)$,

$$\mathbb{P}\left[\frac{\Delta V - \mu_v}{\sigma_v} \leq \frac{\text{VaR}_\alpha - \mu_v}{\sigma_v}\right] = \alpha \text{ et donc } \frac{\text{VaR}_\alpha - \mu_v}{\sigma_v} = -z_\alpha,$$

car $z_{1-\alpha} = -z_\alpha$. Donc pour $0 < \alpha < 1$, la VaR au niveau de confiance α à l'instant $t+1$ est donnée par :

$$\text{VaR}_{\alpha,t+1} = \mu_v - z_\alpha \sigma_v, \alpha \text{ niveau de confiance} \quad (3.5.3)$$

Calcul de la VaR non-linéaire — l'approche par calcul de Gamma

Soient $\{X_i\}_{i=1,\dots,n}$ le rendement de l'instrument non-linéaire et $V(P_1, P_2, \dots, P_n)$ le prix d'un portefeuille composé de n prix d'instruments financiers non-linéaires P_i .

Instrument financier non-linéaire

Nous allons maintenant étudier la variation de deux actifs. Nous commençons par une étude unidimensionnelle pour comprendre la technique du calcul. Soit f une fonction de classe C^∞ . Nous considérons une légère variation h :

$$f(x+h) - f(x) \approx hf'(x).$$

Soit $g(t) = f(x+th) - f(x)$ où $g(0) = 0$ et $g(1) = f(x+h)$. Ainsi, nous avons :

$$\begin{aligned} f(x+h) - f(x) &= g(0) + \int_0^1 g'(t) dt \\ &= \int_0^1 hf'(x+ht) dt \\ &= h \int_0^1 (f'(x+ht) - f'(x) + f'(x)) dt \\ &= h \int_0^1 (f'(x+ht) - f'(x)) dt + hf'(x). \end{aligned}$$

En appliquant à nouveau la formule,

$$\begin{aligned} f(x+h) - f(x) &= h \int_0^1 (f'(x+ht) - f'(x)) dt + hf'(x) \\ &= h \int_0^1 \int_0^1 th(f''(x+hts) - f''(x) + f''(x)) ds dt + hf'(x) \\ &= hf'(x) + h^2 f''(x) \int_0^1 \int_0^1 t ds dt + \int_0^1 \int_0^1 t(f''(x+hts) - f''(x)) ds dt. \end{aligned}$$

Or la quantité $\int_0^1 \int_0^1 t(f''(x+hts) - f''(x)) ds dt$ tend vers zéro lorsque h tend vers 0. Nous pouvons donc négliger ce terme. Ainsi l'expression de la variation de la fonction f devient :

$$f(x+h) - f(x) \approx hf'(x) + \frac{1}{2}f''(x)h^2.$$

3.5. Calcul de la VaR paramétrique

Dans notre cas, nous suivons la même démarche, en multidimensionnel, mais en remplaçant la fonction f par la composition de notre portefeuille $V(P_{1,t}, \dots, P_{n,t})$. Le coefficient h est égal à $\delta_i X_{i,t+1}$. La variation du portefeuille est donc :

$$\Delta V(P_1, \dots, P_n) \simeq \sum_{i=1}^n \delta_i X_{i,t+1} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Gamma_{i,j} X_{i,t+1} X_{j,t+1},$$

où $\Gamma_{i,j} = \frac{\partial^2 V}{\partial P_i \partial P_j} P_i P_j$. La matrice $(\Gamma_{i,j})_{i,j=1,\dots,n}$ est appelée le *gamma du rendement ajusté*. C'est la sensibilité de la valeur du portefeuille lorsque deux composantes varient en même temps. Cette sensibilité décrit la variation de la valeur des actifs lorsqu'une autre donnée varie en même temps d'une quantité $h = \rho P_{i,t}$ où ρ représente le rendement logarithmique journalier des actifs. Le principe est très simple, à chaque fois on fait varier le prix de deux actifs afin d'avoir une idée sur le comportement de ces derniers au cours du temps. Cette évolution est modélisée par une incrémentation à un niveau du prix par un coefficient très petit (qui tends vers zéro mais qui n'est pas nul). Dans notre cas, nous nous arrêtons au second ordre au niveau du développement de la série de Taylor. C'est pourquoi nous nous sommes contentés d'utiliser seulement deux coefficients t et s .

Calcul de la VaR non-linéaire

Afin de pouvoir calculer la VaR non-linéaire, nous utilisons l'écriture matricielle en supposant que le portefeuille contient des instruments financiers non-linéaires. Soient Γ la matrice carrée $n \times n$ des gamma-sensibilités $\Gamma_{i,j}$ et $\Delta V = \delta^T \mathbf{X} + \frac{1}{2} \mathbf{X}^T \Gamma \mathbf{X}$. Nous négligeons les grands termes (nous considérons seulement les quatre premiers moments) et nous supposons que \mathbf{X} suit la loi normale $N(0, \Sigma)$. En revanche, ΔV n'est pas normalement distribuée. Ceci est dû au changement du moment causé par l'asymétrie de la distribution. Donc on ne peut plus travailler sur les quantiles de la distribution normale. Pour cela, nous utilisons la méthode de Cornish-Fisher [LL92] qui consiste à utiliser des moments d'ordre supérieurs à 2 de ΔV . Tout d'abord, nous avons

$$\mathbb{E}[\Delta V] = \mu_{\Delta V} = \frac{1}{2} \text{Tr}[\Gamma \Sigma] \text{ et } \text{Var}(\Delta V) = \sigma_{\Delta V}^2 = \delta^T \Sigma \delta + \frac{1}{2} \text{Tr}[\Gamma \Sigma]^2.$$

Soit Y la variable aléatoire réelle $Y = \frac{\Delta V - \mathbb{E}[\Delta V]}{\sqrt{\text{Var}(\Delta V)}}$. Les moments de Y sont donnés par :

$$\mu_k = \mathbb{E}[Y^k] = \frac{\frac{1}{2} k! \delta^T \Sigma [\Gamma \Sigma]^{k-2} \delta + \frac{1}{2} (k-1)! \text{Tr}[\Gamma \Sigma]^k}{(\delta^T \Sigma \delta + \frac{1}{2} \text{Tr}[\Gamma \Sigma]^2)^{\frac{k}{2}}}, \quad \forall k \geq 3.$$

La quantité $S = \mathbb{E}[Y^3]$, appelée *coefficient d'asymétrie (skewness)*, caractérise l'asymétrie de la loi. Si $Y \sim N(0, 1)$, alors $S = 0$. La quantité $K = \mathbb{E}[Y^4]$, appelée *coefficient d'aplatissement (kurtosis)*, caractérise le pic de la loi. Si $Y \sim N(0, 1)$, alors $K = 3$.

Le quantile de la VaR paramétrique est donné par :

$$z_{\Delta V, \alpha} = z_\alpha + (z_\alpha^2 - 1) \frac{S}{6} + (z_\alpha^3 - 3z_\alpha) \frac{K}{24} - (2z_\alpha^3 - 5z_\alpha) \frac{S^2}{36},$$

où z_α est le α -quantile de la loi normale $N(0, 1)$. Finalement, la VaR au niveau de risque α s'écrit comme suit :

$$\text{VaR}_{\alpha,t+1} = z_{\Delta V, \alpha} \sqrt{\sigma_{\Delta V}^2} + \mu_{\Delta V}. \quad (3.5.4)$$

En pratique, les moments d'ordre 1 à 4 peuvent être estimés empiriquement sur un échantillon de ΔV .

Calcul du gamma

En pratique, lorsque nous manipulons des instruments non-linéaires par exemple les options, nous devons nous baser sur un modèle qui nous permettra de les étudier. Pour ce faire, nous allons adopter le modèle de Black-Scholes-Merton qui repose sur le fait que le prix des instruments financiers non-linéaires (call européen) s'écrit sous la forme suivante :

$$C_i = S_i \phi(d_1^i) - \exp(-rT) K_i \phi(d_2^i), \quad (3.5.5)$$

où

$$d_1^i = \frac{\log\left(\frac{S_i}{K_i}\right) + T\left(r + \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}} \text{ et } d_2^i = \frac{\log\left(\frac{S_i}{K_i}\right) + T\left(r - \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}},$$

où $d_2^i = d_1^i - \sigma_i\sqrt{T}$, S_i est le prix actuel du sous jacent, ϕ est la fonction de répartition de la loi normale centrée réduite, K_i est le prix de l'exercice fixé par l'option dit strike, σ la volatilité du sous jacent, r le taux d'intérêt sans risque et T l'échéance (maturité). Nous avons la parité call-put qui dit que le prix du Put s'écrit sous la forme suivante : $P_i = \exp(-rT) K_i \phi(d_2^i) - S_i \phi(d_1^i)$.

Une petite variation au niveau du prix est la sensibilité δ . En revanche, si cette variation devient forte, on ne peut plus la modéliser en tant que sensibilité mais nous la représentons par une dérivée seconde du prix du sous jacent Γ . Cette modélisation de la variation a pour but de décrire la variation du portefeuille lorsque deux actifs varient en même temps. Le gamma est positif et a la même expression pour le call et le put et nous avons :

$$\Gamma(C_i) = \Gamma(P_i) = \frac{\phi(d_1^i)}{S_i \sigma_i T} > 0, \quad (3.5.6)$$

où Γ est la dérivée seconde par rapport au prix du sous jacent. Étant positif, il décrit la convexité du portefeuille par rapport à ses actifs.

Application numérique

Afin de pouvoir calculer la VaR normale, nous nous basons sur le principe de la fenêtre glissante. Nous estimons les paramètres du modèle gaussien (la moyenne et la variance) à partir des observations des log-rendements (X_1, X_2, \dots, X_T) où T est la dernière observation qui correspond généralement à « aujourd'hui ». La VaR désignera les pertes

3.5. Calcul de la VaR paramétrique

maximales potentielles au temps $T+1$. Pour ce faire, nous utilisons la technique de *fenêtre glissante*. C'est-à-dire que nous fixons un nombre W de jours (ici 2×252 , soit 2 années boursières), et nous estimons les paramètres (moyenne, variance, ...) sur l'échantillon (X_{T-W+1}, \dots, X_T) pour calculer les indicateurs de risque (voir figure 3.5).

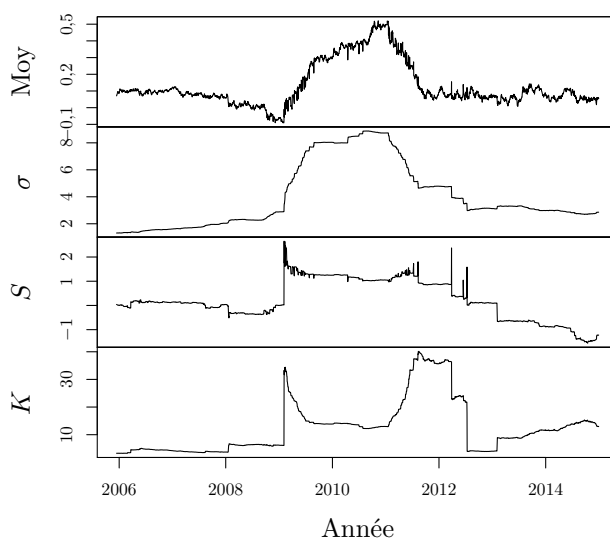


FIGURE 3.5. – Moyenne, écart-type σ , skewness S et kurtosis K calculés sur une fenêtre mobile de deux ans pour le portefeuille introduit au paragraphe 3.3.

La figure 3.6 représente la variation de la Value-at-Risk du modèle normal et avec l'approximation de Cornish-Fisher au cours du temps. Pour cela, nous avons estimé les paramètres du modèle normal (la moyenne et l'écart type) puis nous avons établi la prévision de la VaR.

Nous comparons maintenant la VaR à la valeur de notre portefeuille pour déduire le pourcentage de perte que nous pourrions avoir.

La figure 3.7 montre que la VaR, en période de crise (aux alentours de 2010), présente une fraction qui peut-être importante du prix du portefeuille. Cela veut dire que si les pires scénarios se produisent, alors les pertes peuvent être importantes. De plus, ce graphe démontre une différence importante entre la VaR estimée selon le modèle normal (qui n'utilise que les 2 premiers moments) et l'approximation de Cornish-Fisher (qui utilise les 4 premiers moments).

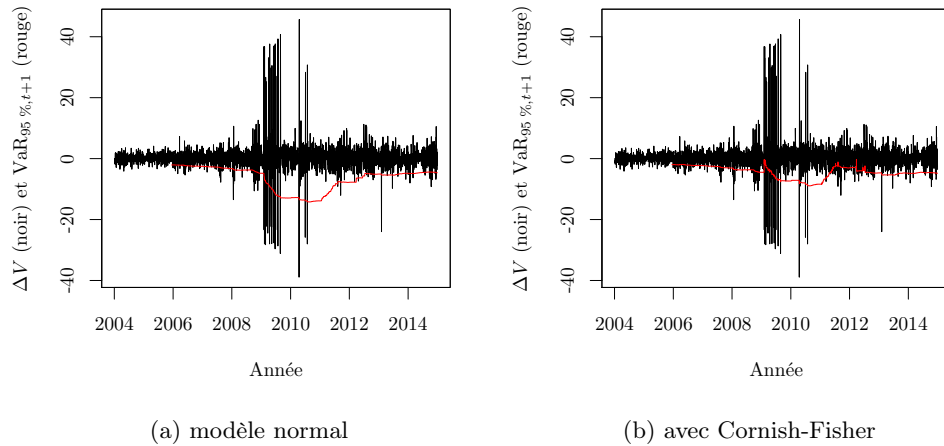


FIGURE 3.6. – L'évolution de la Value-at-Risk quotidienne à un niveau de risque de 5 % suivant le modèle normal (éq. (3.5.2)) ou avec l'approximation de Cornish-Fisher (éq. (3.5.4)).

3.6. Backtesting

Il est toujours préférable de procéder aux testes pour valider la stratégie suivie. Le backtesting est une méthode utilisée afin de s'assurer de la fiabilité de la VaR. Il sert généralement à garantir la performance du modèle de la distribution des rendements financier. En effet, il est essentiel de comparer la théorie et la réalité. Nous nous proposons ainsi de mettre le comportement du modèle en parallèle avec ce qui s'est produit sur le marché boursier. Cette méthode nous est d'une grande utilité pour vérifier l'adéquation du modèle avec la réalité.

Pour ce faire, nous allons relever à partir de nos observations historiques, le nombre de fois où les rendements négatifs (la perte) a dépassé la VaR, et comparer ce chiffre à nos attentes. Ainsi, nous en tirons le nombre de fois où la VaR (les valeurs estimées) a dépassé les valeurs réelles des cours boursiers sur le marché. Si nous obtenons un résultat inférieur à 1 % c'est que notre modèle sous-estime le risque de variation du portefeuille. En revanche, si le pourcentage de dépassement est supérieur à 1 % alors le modèle sur lequel nous nous sommes appuyés sur-estime la fluctuation du marché. Dans le cas où nous avons une parfaite adéquation entre la théorie et la pratique, c'est-à-dire que nous sommes à la couverture, cette comparaison sera égale à 1 %.

Afin de contrôler la gestion des fonds propres, la comité de Bâle III recommande l'utilisation du backtesting. Les gestionnaires de risques doivent relever les dépassements sur une année boursière (soit 252 jours).

En pratique, nous avons testé la VaR avec le modèle normal. Pour ce faire, nous avons

3.6. Backtesting

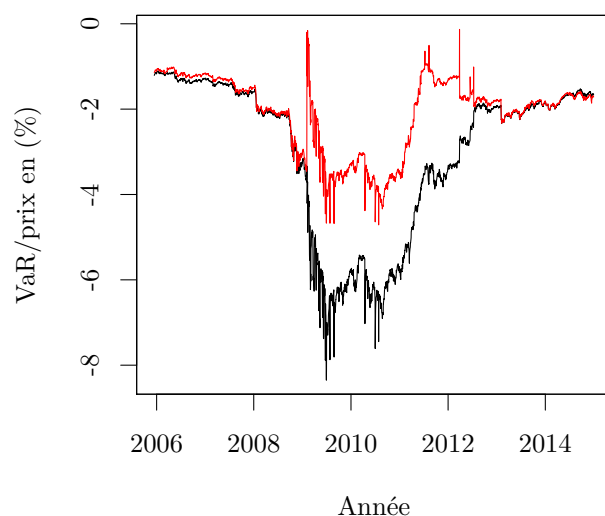


FIGURE 3.7. – Évolution du rapport entre le seuil de perte au niveau de risque $\alpha = 5\%$ et le prix calculés avec le modèle normal (noir) et l'approximation de Cornish-Fisher (rouge).

considéré l'indicateur de violation I_α tel que :

$$I_{(\alpha)} = \begin{cases} 1 & \text{si } X_t \leq \text{VaR}_{\alpha,t+1}(\alpha), \\ 0 & \text{sinon,} \end{cases}$$

où X_t est le rendement financier réel du portefeuille à l'instant t et VaR_α est la perte potentielle maximale du portefeuille. La probabilité de cette violation doit être égale au taux de couverture α . C'est-à-dire :

$$\mathbb{P}[I_{(\alpha)} = 1] = \alpha.$$

Si cette condition n'est pas vérifiée nous avons :

- Si $\mathbb{P}[I_{(\alpha)} = 1] < \alpha$ alors le modèle *sous estime* le risque.
- Si $\mathbb{P}[I_{(\alpha)} = 1] > \alpha$ alors le modèle *sur estime* le risque.

Nous récapitulons dans le tableau 3.1 les résultats du backtesting de la Value-at-Risk lorsque le modèle normal et l'approximation de Cornish-Fisher est utilisée sur nos données.

risque α	1 %	2,5 %	5 %	10 %
Dépassements VaR normale	2,5 %	3,6 %	5,2 %	8,1 %
Dépassements VaR Cornish-Fisher	0,4 %	2,45 %	7,8 %	32,7 %

TABLE 3.1. – Backtesting de la VaR paramétrique (loi normale).

En pratique, l'approximation de Cornish-Fisher ne se révèle pas meilleure que celle par la loi normale en ce qui concerne le calcul de la VaR.

3.7. Calcul de la VaR historique

Dans cette partie, nous nous intéressons au calcul de la VaR historique. Cette méthode repose, d'une part, sur l'hypothèse que les rendements passés peuvent décrire les rendements futurs et d'autre part nous supposons que les rendements financiers X_i sont *i.i.d.* Cette méthode historique repose sur le mécanisme suivant :

1. Nous nous donnons un historique de N observations des prix d'actifs.
2. Nous calculons la variation (gain ou perte) du portefeuille ΔV .
3. Nous ordonnons ces valeurs d'une manière croissante de telle sorte que nous obtenons : $\Delta V^{(1)} \leq \Delta V^{(2)} \leq \dots \leq \Delta V^{(N)}$ avec $\Delta V^{(N)}$ la plus grande valeur de l'échantillon et $\Delta V^{(1)}$ sa plus petite valeur.

Ainsi, la VaR historique au niveau de risque α est notée VaR_α^{HS} et donnée par

$$\text{VaR}_\alpha^{HS} = \Delta V^{([\alpha N])} \text{ où } N \text{ est le nombre d'observations.} \quad (3.7.1)$$

Il s'agit du α -quantile empirique.

Exemple 3.7.1. Nous considérons un portefeuille de N observations et $N = 5000$. Nous souhaitons calculer la VaR à un jour au seuil de 95 %. Nous calculons d'abord la variation du portefeuille. On calcule la différence entre le prix de fermeture à la date $t + 1$ et celui de la date t . Ensuite, nous trions ces rendements suivant un ordre croissant.

$$\text{VaR}_{95\%}^{HS}(\Delta V) = \text{VaR}_{95\%}^{HS}(N(100\% - \alpha\%)) = \text{VaR}_{0,95}^{HS}(5000(100\% - 95\%)).$$

Cela revient à extraire le rendement correspondant au numéro 250 de l'échantillon trié qui désigne la VaR historique.

risque α	1 %	2,5 %	5 %	10 %
VaR normale	-10,0	-8,4	-7,1	-5,5
VaR historique	-8,3	-5,8	-4,3	-3,1

TABLE 3.2. – Estimation de la VaR historique et paramétrique (loi normale) sur l'ensemble des données.

La tableau 3.2 donne la VaR pour quelques niveaux de confiance, calculée avec la VaR historique et la VaR paramétrique sur l'ensemble des données (2818 données). Nous voyons qu'il y a des différences assez significatives entre les niveaux de risque, et que la VaR historique donne une perte moindre par rapport à la VaR paramétrique.

3.8. Calcul de la CVaR paramétrique

Afin de calculer la CVaR paramétrique pour les instruments financiers linéaires, nous supposons que le rendement du portefeuille $\Delta V = \delta^T X$, où X est le vecteur des rendements simples : la différence entre le prix de l'actif d'aujourd'hui et celui d'hier) est normalement distribué avec une moyenne nulle et une matrice de covariance $\delta^T \Sigma \delta$ ($\Delta V \sim N(0, \delta^T \Sigma \delta)$). Nous avons $\text{CVaR}_\alpha = -\mathbb{E}[\Delta V | \Delta V \leq q_\alpha]$ où q_α est le quantile de la distribution normale des rendements. Nous supposons que α est le niveau de risque c'est-à-dire qu'il est plutôt proche de zéro que de 1. Par conséquent le quantile qui lui est associé est négatif et proche de zéro aussi. Ainsi, la moyenne des pertes (les rendements négatifs) qui dépassent la VaR_α est obtenue comme suit :

$$\begin{aligned} \text{CVaR}_\alpha &= -\mathbb{E}[\Delta V | \Delta V \leq q_\alpha] = \frac{-\mathbb{E}[\Delta V 1_{\{\Delta V \leq q_\alpha\}}]}{\mathbb{P}[\Delta V \leq q_\alpha]} = -\frac{1}{\alpha} \mathbb{E}[\Delta V 1_{\{\Delta V \leq q_\alpha\}}] \\ &= -\frac{1}{\alpha \sqrt{\sigma^2 2\pi}} \int_{\mathbb{R}} x \exp\left(\frac{-x^2}{2\sigma^2}\right) 1_{\{x \leq q_\alpha\}} dx = -\frac{1}{\alpha \sqrt{\sigma^2 2\pi}} \int_{-\infty}^{q_\alpha} x \exp\left(\frac{-x^2}{2\sigma^2}\right) dx, \end{aligned}$$

où $\sigma^2 = \delta^T \Sigma \delta$. Nous effectuons le changement de variable $y = x/\sigma$. Comme le quantile z_α de la loi normale standard vérifie $z_\alpha = \frac{q_\alpha}{\sigma}$,

$$\begin{aligned} \text{CVaR}_\alpha &= -\frac{1}{\alpha \sqrt{\sigma^2 2\pi}} \int_{-\infty}^{q_\alpha} x \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = -\frac{1}{\alpha \sqrt{\sigma^2 2\pi}} \int_{-\infty}^{q_\alpha} \sigma y \exp\left(\frac{-y^2}{2}\right) \sigma dy \\ &= \frac{\sigma}{\alpha \sqrt{2\pi}} \exp\left(-\frac{z_\alpha^2}{2}\right). \end{aligned}$$

Nous en déduisons l'expression de la CVaR paramétrique :

$$\text{CVaR}_\alpha = \frac{\sigma}{\alpha} \phi(z_\alpha), \quad (3.8.1)$$

où $\phi(z_\alpha)$ est la densité de la loi normale standard, σ est la volatilité du portefeuille et z_α le quantile de la loi normale standard. Dans le cas où nous souhaitons travailler avec α un seuil de confiance (proche de 1), la démarche demeure la même mais la probabilité sera égale à $(1 - \alpha)$ au lieu de α et le quantile sera $z_{1-\alpha}$. C'est à dire

$$\text{CVaR}_\alpha = \frac{\sigma}{(1 - \alpha)} \phi(z_{1-\alpha}).$$

3.9. Calcul de la CVaR historique

En revanche, nous n'avons pas une expression explicite de la CVaR paramétrique non-linéaire. Afin de l'estimer empiriquement, nous considérons les grandes pertes obtenues de la simulation historique de la VaR et nous en calculons la moyenne.

$$\text{CVaR}_\alpha^{HS} = - \frac{\sum_{i=1}^{[n\alpha]} \Delta V^i}{[n\alpha]}.$$

Le calcul de cette moyenne ne nous rajoute aucune information supplémentaire parce que le fait de supposer que le passé peut décrire le risque futur n'est pas vrai dans la pratique.

risque α	1 %	2,5 %	5 %	10 %
CVaR paramétrique	174,0	69,6	34,7	17,3
CVaR historique	20,8	12,4	8,7	6,2

TABLE 3.3. – Estimation de la CVaR historique et paramétrique.

Le tableau 3.3 représente les estimations de la CVaR suivant la méthode historique et selon la méthode paramétrique du modèle normal. Il est clair que la CVaR historique n'est pas significative vu qu'elle est nettement inférieure à la CVaR paramétrique.

3.10. Inconvénient du modèle normal

La distribution des rendements financiers n'est pas toujours normale. Donc l'utilisation de ce modèle limite la recherche et nous empêche de traiter la queue de la distribution. Donc, le modèle normal n'est pas adapté aux moments de crise et restreint la mesure

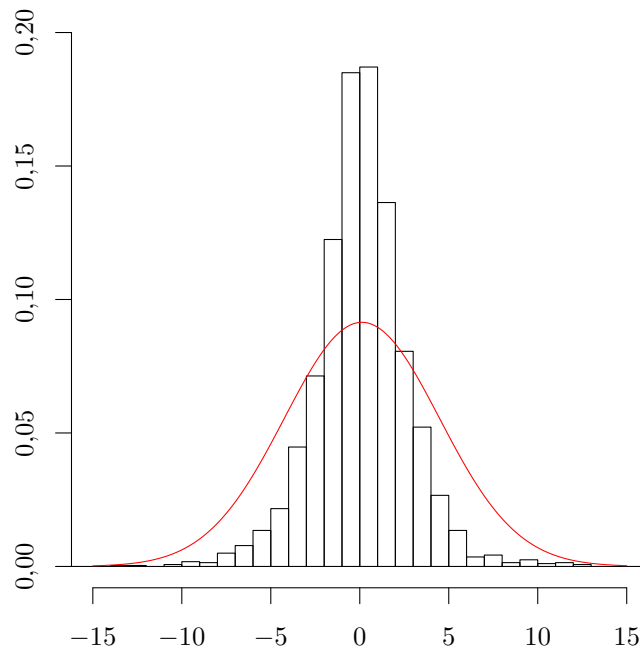


FIGURE 3.8. – Histogramme des log-rendements.

de risque. Cela pose une problématique majeure parce que nous ne pouvons pas estimer correctement la VaR pendant les moments extrêmes.

3.10. Inconvénient du modèle normal

La figure 3.8 représente l'histogramme de la distribution des log-rendements de notre portefeuille introduit au paragraphe 3.3. Nous pouvons voir que la loi normale ne traite pas les valeurs au delà de 0,05. Ce qui confirme l'hypothèse que le modèle normal sous-estime les informations au niveau de la queue. Le calcul de la kurtosis confirme aussi ce fait (voir figure 3.5).

Conclusion

Dans les parties précédentes, nous avons eu recours à la présentation de la VaR et la CVaR et leurs estimations selon le modèle normal. Malgré le fait que cela soit d'une grande importance et d'une grande aide pour les financiers et les gestionnaires, elles présentent certains inconvénients au niveau de la détection du dépassement des pertes au niveau du seuil fixé préalablement aux moments des crises et donc nous ne pouvons pas nous appuyer sur ce modèle. C'est ainsi que nous traiterons la *théorie des valeurs extrêmes* afin de fournir des mécanismes, techniques et des indicateurs performants et rigoureux qui améliorent la performance des portefeuilles.

4. Théorie des valeurs extrêmes (TVE)

Introduction

En dépit de leurs répercussions tragiques sur le marché, les krachs financiers ne sont ni réguliers ni prévisibles. De nos jours, les solutions mathématiques financières accompagnées des techniques numériques, permettent de détecter les événements extrêmes. Dans cette partie nous étudions l'occurrence de survenance des moments de crise. Cela revient à examiner la queue de la distribution des variations de la valeur de notre portefeuille. Il s'agit d'étudier la *théorie des valeurs extrêmes* et déterminer la VaR et la CVaR durant ces moments. La *théorie des extrêmes* se base sur la statistique d'ordre.

Définition 4.0.1 (Statistique d'ordre). Soient $(X_i)_{1 \leq i \leq n}$ des variables aléatoires. On note $(X_{1:n} \leq \dots \leq X_{n:n})$ l'échantillon ordonné qui s'appelle la *statistique d'ordre*. La variable aléatoire $X_{i,n}$ s'appelle la *i^e statistique d'ordre* de l'échantillon.

Remarque 4.0.2. Lorsque n est grand, $X_{i,n}$ est proche du quantile $q_{i/n}$.

La loi des valeurs extrêmes est caractérisée par le théorème de Fisher-Tippet (1928)-Gnedenko (1943) [Lek10].

Théorème 4.0.3 (Fisher-Tippet, Gnedenko). *Sous certaines conditions de régularité sur la fonction de répartition H d'une variable X , il existe $\xi \in \mathbb{R}$ et deux suites $(a_n)_{n \geq 1} > 0$ et $(b_n) \in \mathbb{R}$ telles que pour tout $x \in \mathbb{R}$, on a :*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left[\frac{X_{n,n} - b_n}{a_n} \leq x \right] = H_\xi(x),$$

où $H_\xi(x)$ est une fonction de répartition de loi des valeurs extrêmes. Selon les valeurs de ξ cette loi est soit la loi de Weibull ($\xi < 0$), soit la loi de Gumbel ($\xi = 0$) ou bien la loi de Fréchet ($\xi > 0$).

Dans notre cas, nous nous intéressons seulement à la loi de Fréchet parce qu'elle est la plus significative au niveau du comportement des rendements financiers durant les moments de crise. La loi de Fréchet, associée à la loi de Pareto, est donnée par

$$H_\xi(x) = \begin{cases} 0 & \text{si } x < 0, \\ (e^{-x})^{-\alpha} & \text{si } x \geq 0 \end{cases} \quad \text{avec } \alpha = \frac{1}{\xi} > 0.$$

Dans ce cas, on a $\xi = \frac{1}{\alpha}$, $b_n = 0$ et $a_n = F^{-1} \left(1 - \frac{1}{n} \right)$.

4.1. Loi de Pareto

Selon des études antérieures, il a été prouvé que les rendements des actifs financiers suivent la loi de Pareto pendant les moments de crise. Afin d'estimer l'épaisseur de la queue durant ces moments, nous allons avoir besoin de cette définition.

Définition 4.1.1 (Loi de Pareto). Soit X une variable aléatoire. Nous disons que X suit la loi de Pareto de paramètres (x_0, α) si pour tout $x \geq x_0$, sa fonction de répartition $F_X(x)$ vérifie $F_X(x) = 1 - \left(\frac{x}{x_0}\right)^{-\alpha}$ où $\alpha > 0$ et $x_0 > 0$. La quantité α est l'indice de la queue de la distribution.

Dans le but de simplifier les calculs, on pose $c = x_0^\alpha$. Dans ce cas, l'écriture de la loi de Pareto devient $F_X(x) = 1 - cx^{-\alpha}$ pour tout $\alpha > 0$ et $x > x_0$.

La fonction de densité est donnée par $f(x) = \alpha cx^{-\alpha-1}$ pour tout $\alpha > 0$ et $x > x_0$.

Afin de déterminer la queue de la distribution, nous étudions les moments de la distribution. L'objectif est de déterminer l'épaisseur de la queue et l'ordre à partir duquel l'étude n'est plus significative.

Moment du premier ordre Soit X une variable aléatoire de loi de Pareto de paramètres (x_0, α) . Alors

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x) dx = \int_{\mathbb{R}} \alpha cx^{-\alpha} 1_{[x_0, +\infty[} dx.$$

Proposition 4.1.2. L'existence de ce moment dépend de la valeur de α :

1. Si $\alpha < 1$, alors $\mathbb{E}[X] = +\infty$.
2. Si $\alpha \neq 1$ alors $\mathbb{E}[X] = \frac{\alpha}{\alpha - 1} x_0$.
3. Si $\alpha = 1$ alors $\mathbb{E}[X] = \log(x)$.
4. Si $\alpha > 1$ alors, comme $c = x_0^\alpha$,

$$\mathbb{E}[X] = \alpha c \left[\frac{-1}{(\alpha - 1)x^\alpha - 1} \right]_{x_0}^{+\infty} = \frac{\alpha c}{\alpha - 1} x_0^{1-\alpha} = \frac{\alpha}{\alpha - 1} x_0.$$

Moment d'ordre k Soit X une variable aléatoire de la loi de Pareto de paramètres (x_0, α) . Pour tout $x \in [x_0, +\infty[$,

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k f(x) dx = \int_{\mathbb{R}} \alpha cx^{k-\alpha-1} 1_{[x_0, +\infty[} dx.$$

Proposition 4.1.3. Cette expression dépend de la valeur de $\alpha - k$:

1. Si $\alpha - k = 0$ alors $\mathbb{E}[X^k] = \lim_{y \rightarrow \infty} \alpha c [\log x]_{x_0}^y = +\infty$.
2. Si $\alpha - k > 0$ alors $\mathbb{E}[X^k] = \lim_{y \rightarrow \infty} \alpha c \left[\frac{-1}{(\alpha - k)x^\alpha - k} \right]_{x_0}^y = +\infty$.

4.1. Loi de Pareto

3. Si $\alpha - k < 0$ alors

$$\mathbb{E}[X^k] = \alpha c \left[\frac{-1}{(\alpha - k)x^\alpha - k} \right]_{x_0}^{+\infty} = \frac{\alpha c}{\alpha - k} x_0^{k-\alpha} = \frac{\alpha}{\alpha - k} x_0^k < +\infty.$$

Remarques 4.1.4. Plus α est élevé, plus :

1. la croissance de la fonction de répartition F_X est rapide ;
2. la décroissance de la fonction de densité f est rapide.

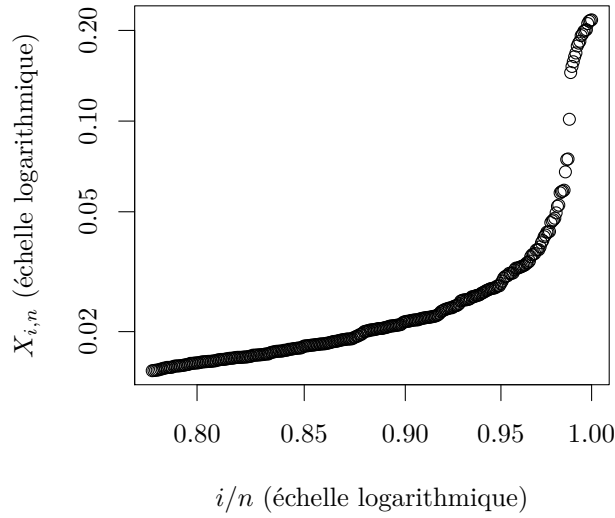


FIGURE 4.1. – Représentation des pertes durant les moments extrêmes.

La figure 4.1 confirme l’hypothèse que nous avons supposée au début de cette partie. En effet, les zones où les points sont alignés sur le graphe montrent que la simulation numérique des log-pertes de la statistique d’ordre de notre échantillon d’observations est adéquate à l’allure de la loi de Pareto (cela se justifie par le fait que $X_{i,n}$ est proche du quantile $q_{i/n}$), avec cependant deux régimes comme nous le verrons, l’un ne concernant que les pertes les plus extrêmes au plus fort de la crise. Localement, dans la zone où les points sont bien alignés à gauche, la pente de cette courbe est bien proche de $\xi = \frac{1}{\alpha}$ où α est le paramètre de la queue de la loi de Pareto que nous avons approché avec l’estimateur de Hill que nous présentons maintenant.

4.1.1. Estimateur de Hill

Nous cherchons à estimer le paramètre α de la loi de Pareto qui décrit l’épaisseur de la queue. Nous utiliserons l’estimateur de Hill qui ne s’applique que pour la loi des extrêmes

dont le paramètre $\xi = \frac{1}{\alpha} > 0$, à savoir la loi de Fréchet. Cet estimateur détermine l'indice de la queue de la distribution des rendements financiers par

$$\widehat{\xi}_{k,n}^H = \frac{1}{k-1} \sum_{j=1}^{k-1} \log \left(\frac{X_{n-j,n}}{X_{n-k,n}} \right).$$

L'estimateur de Hill s'appuie sur la statistique d'ordre, et nous supposons ici que les $X_{j,n}$ sont des réalisations de la statistique d'ordre avec $(X_{n,n} \geq \dots \geq X_{1,n})$. Nous devons choisir le k de sorte que $\frac{k}{n} \rightarrow 0$ avec $k, n \rightarrow +\infty$. C'est-à-dire que k soit assez grand pour traiter les valeurs extrêmes et pas trop petit pour pouvoir aboutir au niveau de stabilité demandé.

k/n	20 %	10 %	5 %	1 %
$\widehat{\xi}$	0,56	0,53	0,52	0,51
$\widehat{\alpha} = 1/\widehat{\xi}$	1,77	1,88	1,91	1,97

TABLE 4.1. – Estimateur de Hill pour les log-pertes.

Le tableau 4.1 indique l'épaisseur de la queue $\widehat{\xi}$ par l'estimateur de Hill pour les plus grandes pertes des log-rendements de notre portefeuille (c'est-à-dire que nous considérons que si X représente les log-rendements, alors nous avons appliqué l'estimateur de Hill à $-X$).

Propriétés 4.1.5. *Nous pouvons énoncer les deux propriétés suivantes qui nous semblent importantes.*

1. $\widehat{\xi}_{k,n}^H$ converge en loi vers ξ lorsque $k, n \rightarrow +\infty$.
2. $\sqrt{k} \frac{\widehat{\xi}_{k,n}^H - \xi}{\xi}$ converge en loi vers $N(0, 1)$ lorsque $k, n \rightarrow +\infty$.

L'estimateur de Hill est l'estimateur du maximum de vraisemblance pour la distribution de Pareto : $\overline{F}(x) = 1 - F(x) = cx^{-1/\xi}$ (voir la définition 4.1.1). Dans le cas général, $\overline{F}(x) = \ell(x)x^{-1/\xi}$ où $\ell(x)$ est une fonction à variation lente. Cette fonction peut être la fonction constante. Les fonctions qui tendent à l'infini vers une constante et la fonction log sont également des fonctions à variation lente au voisinage de l'infini.

Définition 4.1.6 (Fonction à variation lente). On dit qu'une fonction ℓ est à variation lente à l'infini si $\ell(x) > 0$ pour x assez grand et si pour tout $\lambda > 0$ on a :

$$\lim_{x \rightarrow +\infty} \frac{\ell(\lambda x)}{\ell(x)} = 1. \quad (4.1.1)$$

Dans notre cas, la fonction à variation lente est la constante positive $c > 0$ (le paramètre de la loi de Pareto).

Proposition 4.1.7. *Un intervalle de confiance de l'estimateur de Hill est donné par :*

$$I_n(\xi) = \left[\hat{\xi}_n - \frac{z_{\alpha/2} \hat{\xi}_n}{\sqrt{k}}, \hat{\xi}_n + \frac{z_{\alpha/2} \hat{\xi}_n}{\sqrt{k}} \right],$$

où k est la taille de l'échantillon de la statistique d'ordre, $\hat{\xi}_n$ est l'estimateur de Hill du paramètre de la queue de la distribution, $z_{\alpha/2}$ est le $\frac{\alpha}{2}$ -quantile telle que $\mathbb{P}[|N(0,1)| < z_{\alpha/2}] = \alpha$.

Ce résultat a été démontré dans l'article de Cheng et Peng [LP02]. Afin de minimiser les erreurs qui pourront être dues à l'instabilité des estimateurs, à la petite taille des données et à la variance élevée des statistiques d'ordre, nous nous proposons de modéliser l'intervalle de confiance de l'estimateur de Hill afin d'avoir une idée sur sa variation.

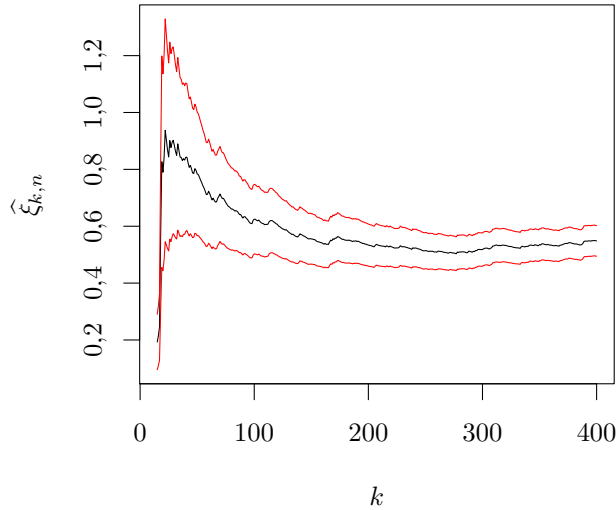


FIGURE 4.2. – Variation de l'estimateur de Hill en fonction de k avec $n = 2818$ observations.

Nous modélisons l'estimateur de Hill $\hat{\xi}$ du paramètre ξ de la queue de la loi de Pareto généralisée (les log-pertes). Nous obtenons le graphe de la figure 4.2 en fonction du seuil k . Cet estimateur semble converger. Nous constatons une légère instabilité due à la petite taille de l'échantillon et à la forte variance due aux statistiques d'ordre des log-pertes.

4.2. Approche des quantiles extrêmes par la méthode des excès

Dans cette partie, nous allons étudier la distribution des pertes. Nous fixons le seuil à partir duquel nous considérons que nos pertes sont extrêmes et nous calculons les indicateurs VaR et la CVaR liés à ces dépassements. Pour ce faire, nous aurons besoin des quantiles extrêmes des excès des pertes. Nous introduisons la Distribution de Pareto Généralisée (GPD) afin d'estimer ces derniers. La théorie antérieure a montré que la distribution des extrêmes se comporte comme la loi de Pareto généralisée à partir d'un certain seuil.

4.2.1. Distribution des excès des pertes

Dans cette partie, nous nous intéressons aux observations des rendements financiers qui dépassent le seuil préalablement fixé ($u < x_F$). Nous allons travailler sur les log-pertes positives $(X_i)_{1 \leq i \leq n}$. Ainsi les k plus grandes valeurs des $(X_i)_{k \leq i \leq n}$ désignent les plus grandes pertes.

Définition 4.2.1. On appelle *point terminal* (*end point*) de la fonction de répartition F le réel x_F défini par $x_F = \sup\{x; F(x) < 1\}$ avec la convention $\sup\{\emptyset\} = \infty$.

A partir de ce point terminal, nous choisissons le seuil u de sorte qu'on ait u très proche de x_F qui tend vers 1. Dans la pratique, nous modélisons les log-pertes positives et nous choisissons le seuil u de sorte que le nombre des log-pertes ne soit pas trop petit pour assurer la stabilité de l'étude et pas trop grand pour ne pas prendre en compte les valeurs qui ne soient pas extrêmes. Nous définissons ainsi la loi des excès des log-pertes qui dépassent le seuil u que nous venons de fixer comme suit.

La figure 4.3 représente la variation des log-pertes qui sont positives. Dans ce cas, nous prenons u de l'ordre de quelques pourcents. Nous fixons le seuil (représenté par la trait rouge dans le graphe) pour étudier les valeurs des log-pertes qui le dépassent.

Définition 4.2.2. Notons Y la variable aléatoire dont la loi est la loi conditionnelle de X sachant $\{X > u\}$ pour un seuil u . Cette variable aléatoire Y traduit l'excès des log-pertes au dessus d'un seuil u . La fonction de répartition d'un excès au dessus du seuil u notée F_u est définie par

$$F_u(y) = \mathbb{P}[X - u \leq y | X > u], \quad \forall y \in [0, x_F - u].$$

Nous réécrivons F_u pour pouvoir étudier le cas où u est grand en

$$F_u(y) = \frac{\mathbb{P}[X - u \leq y; X > u]}{\mathbb{P}[X > u]} = \frac{\mathbb{P}[u < X \leq y + u]}{1 - \mathbb{P}[X \leq u]}.$$

Ainsi nous obtenons pour tout $y \in [0, x_F - u]$ où x_F est le point terminal de la fonction F ,

$$F_u(y) = \frac{F(u + y) - F(u)}{1 - F(u)}. \quad (4.2.1)$$

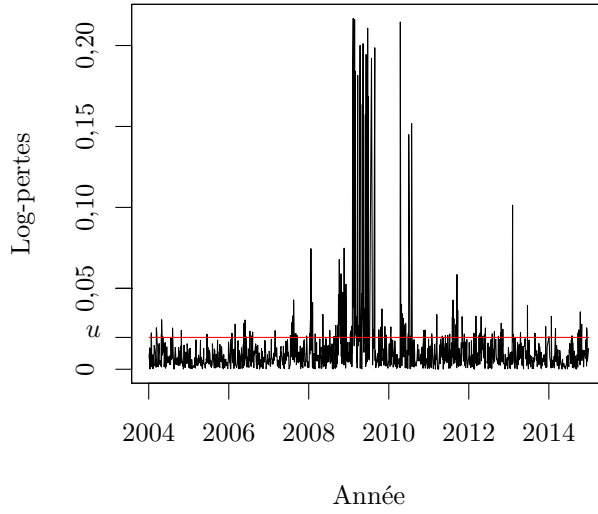


FIGURE 4.3. – Les log-pertes et le seuil $u = 2\%$, ce qui représente 165 des données (soit 5,8% des observations).

L'écriture de la fonction de répartition des excès sous la forme de l'équation (4.2.1) nous permet d'exprimer la fonction des excès F_u en fonction de la fonction de répartition F .

En particulier, lorsque u est très grand (c'est-à-dire qu'il tend vers le point terminal x_F), nous pouvons rapprocher la fonction d'excès F_u à la fonction GPD. Afin d'approcher le quantile extrême, nous utiliserons le théorème de Pickands.

Théorème 4.2.3 (Théorème de Pickands). *Soient F_u la fonction d'excès et $G_{\xi, \beta(u)}$ la fonction de répartition d'une Pareto généralisée présentée plus bas. Nous avons, lorsque u est grand, le résultat suivant :*

$$\lim_{u \rightarrow x_F} \sup_{y \in]0, x_F - u[} |F_u(y) - G_{\xi, \beta(u)}| = 0. \quad (4.2.2)$$

Ce théorème représente la solution de la théorie des valeurs extrêmes car il permet de modéliser le dépassement par la distribution F_u au dessus du seuil u . Ainsi, il suffit simplement d'estimer les paramètres ξ et β et de choisir le seuil u de sorte qu'il soit assez grand pour pouvoir établir la convergence et assez petit afin de laisser des données extrêmes. Nous parlons ici de l'estimation dite *Peaks Over Threshold* (POT).

La figure 4.4 permet de conclure que la distribution des excès des log-pertes au delà du seuil u suit bien une Pareto généralisée. Nous introduisons ainsi la Pareto généralisée.

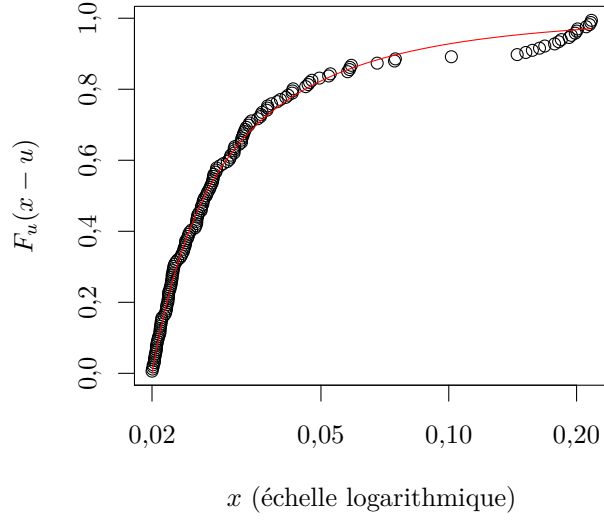


FIGURE 4.4. – La fonction de répartition des excès au delà du seuil $u = 0,02$, et comparaison avec $G_{\xi,\beta(u)}$ où $\xi = 0,97$ et $\beta = 0,007$ sont estimés par un maximum de vraisemblance.

4.2.2. Distribution de Pareto Généralisée (GPD)

Nous considérons une loi de distribution qui présente une queue lourde.

Définition 4.2.4 (Loi de Pareto généralisée). Considérons deux paramètres ξ (le paramètre de la queue) et β (paramètre d'échelle). La fonction de répartition de la distribution de *Pareto généralisée* (GPD) est

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi \frac{x}{\beta})^{-\frac{1}{\xi}} & \text{si } \xi \neq 0, \\ 1 - e^{-x/\beta} & \text{si } \xi = 0, \end{cases} \quad (4.2.3)$$

où $x \geq 0$ si $\xi \geq 0$ et $0 \leq x < -\beta/\xi$ si $\xi < 0$.

Afin de déterminer les quantiles des extrêmes, nous aurons besoin du quantile de la GPD donnée par :

$$G_{\xi,\beta}^{-1}(1 - \alpha) = \begin{cases} \frac{\beta}{\xi}(\alpha^{-\xi} - 1) & \text{si } \xi \neq 0, \\ -\beta \log(\alpha) & \text{si } \xi = 0. \end{cases}$$

4.2. Approche des quantiles extrêmes par la méthode des excès

La densité de la GPD est donnée par :

$$g_{\xi,\beta}(x) = \begin{cases} \frac{1}{\beta} \left(1 + \frac{\xi}{\beta}x\right)^{-1-1/\xi} & \text{si } \xi \neq 0, \\ \frac{1}{\beta} e^{-x/\beta} & \text{si } \xi = 0. \end{cases}$$

Plus la queue est lourde, plus la densité est grande et plus ξ est grand. Par contre plus β est grand plus la partie centrale de la densité s'aplatit. Nous appliquons cette fonction aux log-pertes positifs et nous trouvons :

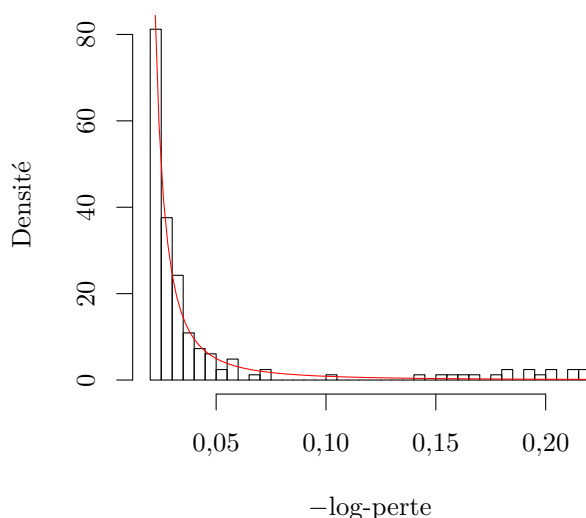


FIGURE 4.5. – Histogramme de log-pertes positives au-dessus du seuil $u = 0,02$ et densité de la loi de Pareto généralisée dont les paramètres sont estimés par un maximum de vraisemblance.

La figure 4.5 est la densité de la Pareto généralisée des log-pertes pour les paramètres $\xi = 0,97$ et $\beta = 0,007$. Ces valeurs sont les estimations des paramètres de la GPD par la méthode du maximum de vraisemblance.

Le moment d'ordre k de la GPD Soit X une variable aléatoire qui suit la GPD. On a :

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k g_{\xi,\beta}(x) dx = \frac{1}{\beta} \int_{\mathbb{R}} \frac{x^k}{\left(1 + \frac{\xi}{\beta}x\right)^{1+1/\xi}} dx.$$

Or au voisinage de l'infini, nous avons : $\frac{x^k}{\left(1 + \frac{\xi}{\beta}x\right)^{1+1/\xi}} \sim \frac{\beta}{\xi} x^{k-1/\xi-1}$. Ainsi

1. Si $k - 1 - 1/\xi < 1$ alors $\mathbb{E}[X^k] < \infty$.
2. Si $k - 1 - 1/\xi > 1$ alors $\mathbb{E}[X^k] \rightarrow \infty$.

Nous en déduisons que le k^e moment d'une variable aléatoire qui suit la GPD est fini si $\xi < 1/k$.

4.2.3. Estimation de la VaR et la CVaR

Nous pouvons réécrire l'expression de la distribution des excès F en fonction de la GPD. D'après le théorème 4.2.3, nous avons :

$$G_{\xi, \beta(u)}(y) = F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

Cela implique $(1 - F(u))G_{\xi, \beta(u)}(x - u) = F(x) - F(u)$ car $y = x - u$. Donc pour u grand,

$$F(x) = [1 - F(u)]G_{\xi, \beta(u)}(x - u) + F(u), \quad \forall x > u.$$

Afin de déterminer le quantile correspondant au seuil u nous estimons F par la fonction empirique $\hat{F}(u) = \frac{n - N_u}{n}$ où n est le nombre d'observations des log-pertes positifs et N_u est le nombre de dépassements par rapport au seuil choisi u . Ainsi nous avons :

$$\hat{F}(x) = 1 - \frac{N_u}{n} \left(1 + \frac{\hat{\xi}}{\hat{\beta}}(x - u)\right)^{-1/\hat{\xi}}; \quad \forall x > u, \quad (4.2.4)$$

où $\hat{\xi}$ est l'estimateur du paramètre de la queue ξ et $\hat{\beta}$ l'estimateur du paramètre d'échelle β .

Cela signifie qu'à une probabilité $(1 - \alpha) > F(u)$, nous pouvons associer l'estimation de la perte maximale potentielle donnée par l'indicateur VaR à partir du quantile de la distribution des pertes \hat{F} . Nous avons $\hat{F}(x) = 1 - \alpha$ pour tout $x > u$. Nous remplaçons \hat{F} par son expression

$$1 - \alpha = 1 - \frac{N_u}{n} \left(1 + \frac{\hat{\xi}}{\hat{\beta}}(x - u)\right)^{-1/\hat{\xi}}, \quad \forall x > u.$$

Nous obtenons

$$1 + \frac{\hat{\xi}}{\hat{\beta}}(x - u) = \left(\alpha \frac{n}{N_u}\right)^{-\xi}, \quad \forall x > u.$$

Par la suite $x - u = \left[\left(\alpha \frac{n}{N_u}\right)^{-\xi} - 1\right] \frac{\hat{\beta}}{\hat{\xi}}$ pour tout $x > u$.

4.2. Approche des quantiles extrêmes par la méthode des excès

Nous pouvons conclure ainsi que, lorsque X est un log-rendement (et non une log-perte positive),

$$\text{VaR}_\alpha^+(X) = \left[\left(\alpha \frac{n}{N_u} \right)^{-\xi} - 1 \right] \frac{\hat{\beta}}{\hat{\xi}} + u, \quad \forall x > u. \quad (4.2.5)$$

	$\hat{\xi} = 0,97 \quad \hat{\beta} = 0,007$			
Seuil α	1 %	2,5 %	5 %	10 %
VaR log-rendements	-59,1 %	-25,3 %	-13,7 %	-7,6 %
VaR prix	-124,6	-62,5	-35,71	-20,63

TABLE 4.2. – Estimation des paramètres de la GPD et de la VaR quotidienne calculée sur l'ensemble des données.

Le tableau 4.2 donne les valeurs du paramètre d'échelle et l'épaisseur de la queue. Nous avons également estimé la Value-at-Risk suivant le modèle de la Pareto sur l'ensemble des données. Pour cela, nous avons tout d'abord calculé la VaR sur les log-rendements, en supposant qu'ils suivent le modèle de Pareto. Nous avons ensuite calculé la VaR_{prix} au niveau du prix grâce à la formule $\text{VaR}_{\text{prix}} = P(\exp(\text{VaR}) - 1)$, où P est la valeur terminale du prix (ici $P = 279\text{€}$). Notons que contrairement à la VaR calculée selon le modèle normal, les pertes peuvent être très importantes. Le tableau 3.2 donne pour le modèle normal une perte de 10€ avec un risque de 1 %. Avec le modèle de Pareto, la perte est de 125€, c'est-à-dire environ la moitié de la valeur finale du portefeuille.

Lorsque les pertes dépassent la VaR, nous passons au calcul de la CVaR. Nous avons :

$$\text{CVaR}_\alpha = \mathbb{E}[X|X > \text{VaR}_\alpha^+] = \mathbb{E}[X - \text{VaR}_\alpha^+ + \text{VaR}_\alpha^+ | X > \text{VaR}_\alpha^+].$$

Nous pouvons conclure que :

$$\text{CVaR}_\alpha = \text{VaR}_\alpha^+ + \mathbb{E}[X - \text{VaR}_\alpha^+ | X > \text{VaR}_\alpha^+]. \quad (4.2.6)$$

Nous remarquons que la CVaR_α est la somme de la VaR_α^+ et la moyenne de la distribution des excès $F_{\text{VaR}_\alpha^+}(y)$. Cette moyenne est appelée *la fonction moyenne d'excès* de la VaR_α^+ . Nous considérons la VaR_α^+ comme étant le seuil à partir duquel nous allons déterminer le taux de dépassement. Nous avons dans ce cas l'expression de la moyenne d'une variable aléatoire qui suit la GPD :

$$F_{\text{VaR}_\alpha}(y) = G_{\xi, \beta + \xi \text{VaR}_\alpha^+}(y).$$

Calcul de la moyenne de l'excès d'une variable aléatoire qui suit la GPD Soit $X - u$ une variable aléatoire telle que X suit la GPD. On sait d'après le calcul antérieur que la moyenne d'ordre k existe et est finie si et seulement si $k < 1/\xi$. Donc, pour $k = 1$ la condition d'existence est vérifiée lorsque $\xi < 1$. Par conséquent, pour tout $x \geq 0$,

$$e(u) = \mathbb{E}[X - u | X > u] = \frac{\mathbb{E}[(X - u)1_{\{X > u\}}]}{\mathbb{P}[X > u]}.$$

Définition 4.2.5. La fonction e s'appelle la fonction *moyenne des excès* (*mean excess*).

Nous commençons par calculer l'espérance :

$$\begin{aligned}\mathbb{E}[(X - u)1_{\{X > u\}}] &= \int_{\mathbb{R}} (x - u)1_{\{x > u\}} g_{\xi, \beta}(x) dx \\ &= \frac{1}{\beta} \int_u^{+\infty} (x - u) \left(1 + \frac{\xi}{\beta} x\right)^{-1-1/\xi} dx.\end{aligned}$$

Nous faisons le changement de variable $y = x - u$:

$$\mathbb{E}[(X - u)1_{\{X > u\}}] = \frac{1}{\beta} \int_0^{+\infty} y \left(1 + \frac{\xi}{\beta} (y + u)\right)^{-1-1/\xi} dy.$$

Nous procédons à une intégration par partie :

$$\mathbb{E}[(X - u)1_{\{X > u\}}] = \left[-y \left(1 + \frac{\xi}{\beta} (y + u)\right)^{-1/\xi} \right]_0^{+\infty} + \frac{1}{\beta} \int_0^{+\infty} \left(1 + \frac{\xi}{\beta} (y + u)\right)^{-1/\xi} dy.$$

Le premier terme s'annule parce que $\xi < 1$. Il nous reste à intégrer la deuxième partie :

$$\begin{aligned}\mathbb{E}[(X - u)1_{\{X > u\}}] &= \frac{1}{\beta} \int_0^{+\infty} \left(1 + \frac{\xi}{\beta} (y + u)\right)^{-1/\xi} dy \\ &= \left[\left(1 + \frac{\xi}{\beta} (y + u)\right)^{1-1/\xi} \right]_0^{+\infty} = \frac{-\beta}{\xi - 1} \left(1 + \frac{\xi}{\beta} u\right)^{1-1/\xi}.\end{aligned}$$

Maintenant, nous calculons la probabilité que les rendements X dépassent le seuil u :

$$\mathbb{P}[X > u] = 1 - \mathbb{P}[X \leq u] = 1 - \int_{-\infty}^u \frac{1}{\beta} \left(1 + \frac{\xi}{\beta} x\right)^{-1-1/\xi} dx.$$

Or pour le modèle de la *Pareto généralisée*, les valeurs de x sont toujours positives. Donc,

$$\begin{aligned}\mathbb{P}[X > u] &= 1 - \int_0^u \frac{1}{\beta} \left(1 + \frac{\xi}{\beta} x\right)^{-1-1/\xi} dx \\ &= 1 - \left[\left(1 + \frac{\xi}{\beta} x\right)^{1/\xi} \right]_0^u = 1 - \frac{-\xi\beta}{\xi\beta} \left[\left(1 + \frac{\xi}{\beta} x\right)^{-1/\xi} \right]_0^u \\ &= 1 + \left[\left(1 + \frac{\xi}{\beta} u\right)^{-1/\xi} - 1 \right] = \left(1 + \frac{\xi}{\beta} u\right)^{-1/\xi}.\end{aligned}$$

Donc, pour X une variable aléatoire qui suit la GPD, la moyenne d'excès au seuil u est

$$e(u) = \mathbb{E}[X - u | X > u] = \frac{\beta}{1 - \xi} \left(1 + \frac{\xi}{\beta} u\right). \quad (4.2.7)$$

Dans notre cas, nous retenons le seuil u de manière que la moyenne des excès soit approximativement linéaire.

4.2. Approche des quantiles extrêmes par la méthode des excès

Remarque 4.2.6. Plus la queue de la distribution est épaisse (c'est-à-dire pour $\xi > 0$ est grand), plus la fonction de la moyenne des excès tend rapidement vers l'infini.

Nous définissons également la fonction de moyenne des excès de la VaR_α^+ qui dépasse le seuil u . Soit Y une variable aléatoire telle que $Y = [X - u | X > u]$. Nous avons l'égalité en loi suivante :

$$\text{Loi}[X - u | X > u] = \text{Loi}[Y - (\text{VaR}_\alpha^+ - u) | Y > \text{VaR}_\alpha^+ - u].$$

Donc nous pouvons écrire à présent la moyenne des excès associée à la VaR^+ en remplaçant le u par $\text{VaR}_\alpha^+ - u$ dans l'équation (4.2.7) :

$$e(\text{VaR}_\alpha) = \mathbb{E}[X - \text{VaR}_\alpha^+ | X > \text{VaR}_\alpha^+] = \frac{\beta + \xi(\text{VaR}_\alpha^+ - u)}{1 - \xi}. \quad (4.2.8)$$

Calculons la moyenne lorsque les log-pertes $(X_i)_{1 \leq i \leq n}$ dépassent la $\text{VaR}_\alpha^+ > u$. Nous nous basons sur l'expression de la CVaR de l'équation (4.2.6) et nous remplaçons la moyenne par le résultat trouvé dans l'équation (4.2.8). Nous obtenons ainsi :

$$\text{CVaR}_\alpha = \frac{\text{VaR}_\alpha^+ + \beta - \xi u}{1 - \xi}. \quad (4.2.9)$$

α	1 %	2,5 %	5 %	10 %
CVaR_α	13,5	5,6	2,9	1,5

TABLE 4.3. – Estimation de la CVaR_α pour un seuil $u = 0,02$.

Le tableau 4.3 indique l'estimation de la Conditional Value-at-Risk suivant le modèle de la Pareto généralisée. C'est la moyenne des pertes si jamais ces dernières dépassent la VaR^+ .

4.2.4. Maximum de vraisemblance (MLE)

Dans cette partie, nous souhaitons estimer les paramètres ξ et β de la GPD. Nous aurons recours à la fonction du maximum de vraisemblance (*maximum likelihood estimation*, MLE) pour estimer ces paramètres. Cette méthode présente un avantage parce qu'il est plus facile d'avoir un échantillon d'excès. Dans la pratique, nous choisissons tout d'abord le seuil u de la fonction Mean Excess Plot. Ensuite, nous sélectionnons les observations au delà du seuil u (nous travaillons sur la statistique d'ordre) et nous appliquons la GPD aux excès des log-rendements positifs. Enfin, la MLE sélectionne les estimateurs $\hat{\xi}$ et $\hat{\beta}$ de sorte qu'ils maximisent la fonction de vraisemblance $L(\hat{\xi}, \hat{\beta} | y) = \max_{\xi, \beta} L(\xi, \beta | y)$

où $L(\hat{\xi}, \hat{\beta} | y) = \max_{\xi, \beta} \prod_{i=1}^n g_{\xi, \beta}(y_i)$ avec $y = [y_1, \dots, y_n]$ notre échantillon d'observation. Cela revient à maximiser la fonction log-vraisemblance définie par :

$$\ell(\hat{\xi}, \hat{\beta} | y) = \max_{\xi, \beta} \log L(\xi, \beta | y) = \max_{\xi, \beta} \sum_{i=1}^n \log(g_{\xi, \beta}(y_i)).$$

Nous écrivons la fonction de vraisemblance d'une manière explicite à partir de la fonction de densité de la Pareto généralisée.

$$\ell(\xi, \beta|y) = \begin{cases} -n \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log(g_{\xi, \beta}(y_i)) & \text{si } \xi \neq 0, \\ -n \log \beta - \frac{1}{\beta} - \frac{1}{\beta} \sum_{i=1}^n y_i & \text{si } \xi = 0. \end{cases} \quad (4.2.10)$$

Afin d'estimer les paramètres ξ et β , nous allons dériver l'équation (4.2.10) par rapport à β . Nous obtenons ainsi les estimations suivantes :

$$\frac{\partial}{\partial \beta} \left[-n \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log\left(1 + \frac{\xi}{\beta} y_i\right) \right] = 0$$

soit

$$\frac{-n}{\beta} - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \frac{\frac{-\xi y_i}{\beta^2}}{1 + \frac{\xi}{\beta} y_i} = 0$$

soit

$$\frac{-n}{\beta} + \frac{\xi}{\beta^2} \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \frac{y_i}{1 + \frac{\xi}{\beta} y_i} = 0$$

et donc

$$\frac{1 + \xi}{\beta} \sum_{i=1}^n \frac{y_i}{1 + \frac{\xi}{\beta} y_i} = n.$$

Nous obtenons l'estimation de β qui est donnée par :

$$\hat{\beta} = \frac{1 + \hat{\xi}}{n} \sum_{i=1}^n y_i \left(1 + \frac{\hat{\xi}}{\hat{\beta}} y_i\right)^{-1}.$$

Maintenant nous dérivons l'équation 4.2.10 par rapport à ξ :

$$\frac{\partial}{\partial \xi} \left[-n \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log\left(1 + \frac{\xi}{\beta} y_i\right) \right] = 0$$

soit

$$\frac{1}{\xi^2} \sum_{i=1}^n \log\left(1 + \frac{\xi}{\beta} y_i\right) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \frac{\frac{y_i}{\beta}}{\left(1 + \frac{\xi}{\beta} y_i\right)} = 0$$

ou encore

$$\frac{1}{\xi^2} \sum_{i=1}^n \log\left(1 + \frac{\xi}{\beta} y_i\right) - \frac{\xi + 1}{\xi \beta} \sum_{i=1}^n y_i \left(1 + \frac{\xi}{\beta} y_i\right)^{-1} = 0.$$

4.2. Approche des quantiles extrêmes par la méthode des excès

Or $\sum_{i=1}^n y_i \left(1 + \frac{\xi}{\beta} y_i\right)^{-1} = \frac{n\beta}{\xi + 1}$. En remplaçant cette somme dans l'équation,

$$\frac{1}{\xi} \sum_{i=1}^n \log \left(1 + \frac{\xi}{\beta} y_i\right) - \frac{\xi + 1}{\beta} \frac{n\beta}{\xi + 1} = 0,$$

d'où finalement

$$\frac{1}{\xi} \sum_{i=1}^n \log \left(1 + \frac{\xi}{\beta} y_i\right) = n.$$

Nous obtenons ainsi l'estimation de ξ qui est donnée par :

$$\hat{\xi} = \frac{1}{n} \sum_{i=1}^n \log \left(1 + \frac{\hat{\xi}}{\hat{\beta}} y_i\right).$$

Nous remarquons que $\hat{\beta}$ et $\hat{\xi}$ dépendent de $\frac{\xi}{\beta}$. Nous posons $\theta = \frac{\xi}{\beta}$,

$$\hat{\beta} = \frac{1 + \hat{\xi}}{n} \sum_{i=1}^n y_i \left(1 + \frac{\hat{\xi}}{\hat{\beta}} y_i\right)^{-1}, \text{ soit } \frac{\hat{\beta}}{\hat{\xi}} = \frac{\hat{\xi} + 1}{\hat{\xi} n} \sum_{i=1}^n y_i \left(1 + \frac{\hat{\xi}}{\hat{\beta}} y_i\right)^{-1}.$$

L'estimation de θ est donnée par :

$$\frac{1}{\hat{\theta}} = \frac{1}{n} \left(1 + \frac{1}{\hat{\xi}}\right) \sum_{i=1}^n \frac{y_i}{1 + \hat{\theta} y_i}.$$

Nous pouvons résoudre cette équation numériquement grâce à la technique itérative. Il suffit de disposer d'une valeur initiale pour aboutir à la solution.

En pratique, nous modélisons le graphe de la fonction des excès des pertes au delà du seuil u afin de fixer la valeur de ce dernier. Il s'agit de considérer les log-rendements positifs qui reflètent les grosses pertes. Par la suite, nous ordonnons cet échantillon puis nous en sélectionnons les k plus grandes observations. Nous leur appliquons la distribution des excès qui est égale à la GPD selon le théorème 4.2.3. Enfin nous dessinons cette distribution. Nous pouvons à ce stade choisir graphiquement le seuil u .

Le *Mean Excess Plot* est un outil spécifique pour retenir le seuil performant de la queue (voir figure 4.6). Si les log-pertes positives suivent une loi de Pareto avec $0 < \xi < 1$, alors la formule (4.2.7) prédit que $e(u)$ se comporte linéairement avec u . La forme du graphe dans la figure 4.6 justifie notre choix de prendre $u = 0,02$. Pour $u > 0,1$, nous ne disposons que de 18 données (sur 2817). L'estimateur de maximum de vraisemblance donne alors $\hat{\xi} = -2,88$ et $\hat{\beta} = 0,36$. Cela veut dire que les log-pertes sont bornées.

4.2.5. QQ-Plot

La QQ-plot est le graphe du quantile-quantile, c'est une technique graphique qui sert à déterminer si un échantillon suit une certaine distribution. Nous avons eu recours à cette

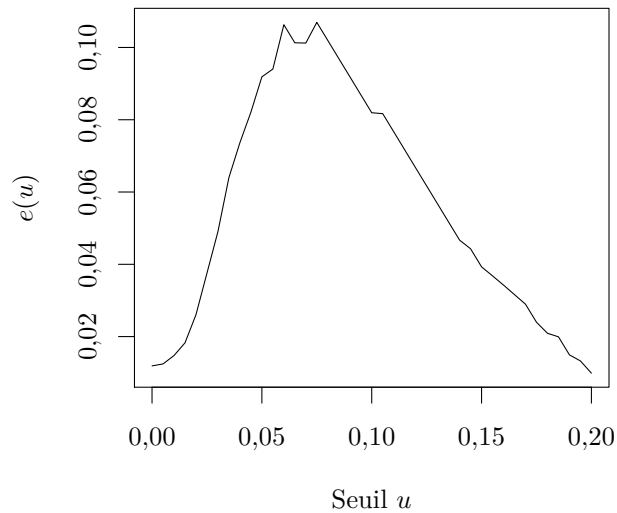


FIGURE 4.6. – Graphe de la moyenne des excès.

méthode pour nous assurer que la queue de notre distribution (les log-pertes positifs qui désignent les plus grandes pertes) suit bien la loi de Pareto généralisée. Cette méthode repose sur le calcul des quantiles de la queue, les tracer sur un plan et les comparer à la droite $y = x$. Pour un échantillon en parfaite adéquation avec la loi, ces points qui représentent les quantiles suivent cette droite. La taille réduite de la queue de notre distribution risque d'élargir l'intervalle d'erreur. Dans le cas où $u = 0,02$, la fraction la plus petite des données colle bien à la GPD, puis s'en éloigne. Dans le cas $u = 0,10$, les quelques données les plus grandes collent bien à la GPD, mais cela correspond aux risques les plus extrêmes (c'est-à-dire plus de 10 % de perte en une seule journée) qui ne sont observés qu'aux instants les plus forts de la crise autour de 2008.

Ces graphes confirment à la fois la pertinence d'utiliser la GPD, mais aussi de la difficulté d'étalonner les paramètres, car il y a manifestement ici deux régimes différents. Cela était déjà visible dans le graphe 4.6.

Conclusion

Dans ce chapitre, nous avons étudié les valeurs extrêmes. Nous avons mis en évidence que les rendements financiers (la queue) suivent la loi de Pareto généralisée pendant les moments de crise. Nous avons considéré que les rendements sont constants (indépendants du temps). Empiriquement ce n'est pas toujours le cas parce que pendant les moments où la volatilité est élevée, les résultats seront erronés.

4.2. Approche des quantiles extrêmes par la méthode des excès

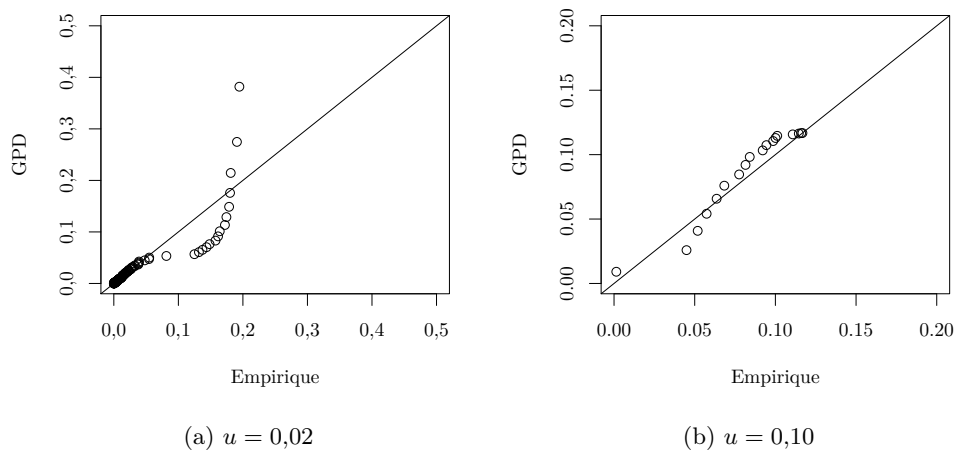


FIGURE 4.7. – QQ-plot de la queue de distribution empirique contre la GPD.

Conclusion générale

Notre projet a été conçu en deux grandes parties : La première est d'étudier le risque financier d'un portefeuille. Durant cette étape nous avons calculé la perte maximale potentielle qu'un portefeuille est susceptible d'avoir. Nous avons également pris en considération le dépassement de cette dernière et donc nous avons estimé la moyenne de ce dépassement afin de minimiser les pertes des investisseurs. Nous avons modélisé la VaR avec le modèle normal et avons constaté que ce dernier n'est pas valable durant les moments de krach vu que ce dernier ne traite pas la queue de la distribution. Nous avons par conséquent pris en compte les moments de crise où le marché financier est instable et les variations des rendements présentent des pics importants. Nous avons commencé par étudier l'épaisseur de la queue de la distribution. Nous l'avons estimée par l'estimateur de Hill. Nous avons également étudié son comportement durant les moments de crise. Par la suite, nous sommes passés à l'implémentation d'une solution numérique de ces études et nous avons pensé à aider les décideurs à effectuer leurs analyse. Ainsi, nous avons réalisé un outil décisionnel qui permet d'automatiser le processus de supervision des cours boursiers, d'augmenter la fiabilité de ce dernier, assurer le gain de temps et d'approfondir les analyses qui concernent ce processus.

L'application réalisée permet aux décideurs d'avoir une vision claire sur l'évolution du marché boursier. Elle offre également une vision globale qui peut être spécifiée selon les axes d'analyses prédéfinis. De plus, les analyses de données (Série temporelle) nous ont permis d'approfondir l'étude afin d'extraire des informations et formuler des conclusions ainsi que des recommandations pour aider les décideurs à minimiser le risque de pertes importantes.

Nous tenons à souligner également que ce projet a été l'occasion d'acquérir une expérience très enrichissante en travaillant sur des aspects à la fois académiques et professionnels dans le secteur des mathématiques financières et du business intelligence. Il est également fondamental de préciser la difficulté que nous avons eu au début du projet. Nous avons travaillé sur de nouvelles notions en finance, en mathématiques et sur des nouvelles plateformes ce qui a enrichi nos connaissances mais aussi a constitué un défi pour réussir le projet.

Grâce au bon encadrement, le suivi, la bonne gestion de temps et à la méthodologie Agile, nous sommes arrivés à bien surmonter tous les problèmes que nous avons eu.

Durant notre travail, nous avons supposé la volatilité constante au cours du temps. Ceci n'est pas vrai parce que la fluctuation des cours des actifs affecte les rendements financiers et donc affectent la volatilité. Nous envisageons mener une étude sur cela ainsi que sur l'étude de la corrélation entre les actifs qui augmente le risque.

A. Portail décisionnel

Introduction

Dans ce chapitre, nous nous proposons de détailler la mise en place d'un portail décisionnel ainsi que toutes les étapes suivies pour réaliser cette partie. En premier lieu, nous expliquons les différentes problématiques du sujet et ensuite nous décrivons la méthodologie de développement que nous avons adoptée durant cette étape.

A.1. Présentation de l'application

A.1.1. Contexte de l'application

Cette partie consiste à présenter l'outil décisionnel que nous avons développé. L'objectif est d'aider les gestionnaires de risques et les économistes à optimiser leurs prises de décision sur la variation des cours boursiers et à réaliser une étude prédictive sur la fluctuation de ces derniers et leurs impacts sur le marché financier.

D'une part, cette application donne une vision claire et approfondie sur le processus de variation des cours à travers des tableaux de bords complets qui assureront une interaction instantanée avec les données.

D'autre part, il porte sur l'analyse des données afin de faciliter la classification des entreprises et en offrant une prédiction de fluctuation des prix des actifs.

A.1.2. Démarche de réalisation de l'application BI

L'informatique décisionnelle désigne les moyens, les outils et les techniques qui permettent de collecter, consolider, modéliser et restituer les données immatérielles d'une entreprise. Elle permet d'automatiser la tâche d'analyse de la situation actuelle en offrant une vue globale et transversale des données qui proviennent des différentes sources et d'optimiser le mécanisme de prise des décisions qui affecteront la performance et la rentabilité de la société. La chaîne décisionnelle qui apparait dans la figure A.1 comprend les différentes phases suivantes :

1. La phase de modélisation : Cette phase est le moment où les données sont stockées sous une forme adaptée pour les analyses que nous souhaitons effectuer. Elle contient notamment le Data Warehouse chargé de centraliser les données. Elle fait aussi intervenir les notions de cubes et de Datamarts nécessaires pour coller au plus près des attentes métiers.

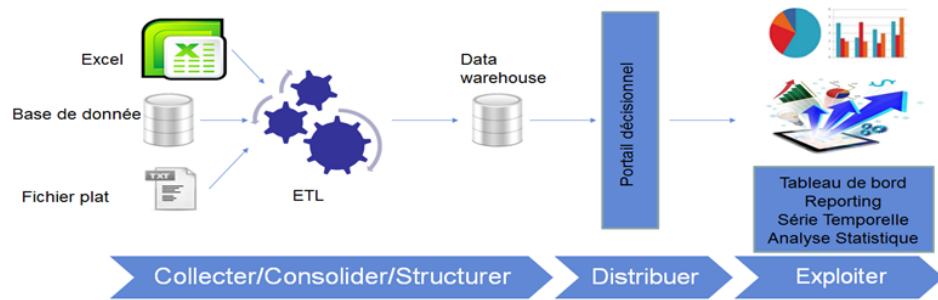


FIGURE A.1. – Le processus BI.

2. La phase d'alimentation : C'est au niveau de cette phase qu'on va alimenter notre Data Warehouse à travers des processus ETL qui se chargeront de récupérer toutes les données nécessaires depuis les différentes sources de stockage.
3. La phase de restitution : C'est au niveau de cette phase que les différents outils de restitution vont intervenir. Nous trouverons des outils de reporting, des portails d'accès à des tableaux de bord, des outils de navigation dans des cubes, ou des outils de statistiques.
4. La phase d'analyse : C'est au niveau de cette phase que les utilisateurs finaux interviennent et analysent les informations qui leur sont fournies. Elle peut aussi faire intervenir des spécialistes en analyse pour utiliser des outils de statistique et ressortir des prévisions ou des estimations futures (Série temporelle).

A.1.3. Objectif de l'application

Notre application se compose de deux parties :

1. Une application Java qui consiste à automatiser la tâche d'extraction des données à partir d'un site internet Boursorama ce qui va nous permettant d'offrir une meilleure visibilité sur la variation des actions dans la bourse ainsi qu'une meilleure crédibilité des analyses (de données, statistiques, financière, ...) à travers les tableaux de bords.
2. Une autre BI, nous allons mettre en place un système décisionnel qui permet l'analyse décisionnelle et prédictive à travers des tableaux de bords interactifs et complets.

A.2. Réalisation de l'application

A.2.1. Extraction des données à partir du web

Pour assurer la mise à jour de notre base d'analyse d'une façon quotidienne, nous avons pensé à automatiser cette tâche en l'alimentant par des données à partir du site web. Pour ce faire, on a implémenté une application Java en utilisant la bibliothèque JSOUP qui nous a permis d'avoir les données nécessaires pour actualiser nos analyses

A.2. Réalisation de l'application

d'une manière instantanée. Nous avons choisi d'extraire les données à partir du site



FIGURE A.2. – Extraction des données à partir du site web.

"Boursorama Banque" qui représente un portail d'information économique et financière fiable. Ce site appartient à la Société générale. C'est l'unique banque en ligne certifiée AFNOR qui offre une mise à jour des données de la Bourse de Paris et des indices Euronext en temps réel. Afin de pouvoir capter le flux de données à partir du site, nous avons développé une application Java se basant sur le parseur HTML, JSOUP. Ces données seront stockées dans une base de données intermédiaire (ORACLE) dont nous allons nous en servir pour alimenter notre entrepôt de données (Data Warehouse). Ces étapes d'extraction sont expliquées dans la figure A.2.

Outil technique utilisé Durant cette étape du projet, nous avons utilisé Eclipse comme outil de travail. C'est un projet de la Fondation Eclipse mettant à la disposition des développeurs un environnement de production de logiciels s'appuyant principalement sur Java.

A.2.2. Modélisation du Data Warehouse

Définition du Data Warehouse L'entrepôt de données « Data Warehouse », est une base de données dédiée au stockage de l'ensemble des données utilisées dans le cadre de la prise de décision et de l'analyse décisionnelle. Le Data Warehouse est exclusivement réservé à cet usage. Il est alimenté par des données depuis les bases de production grâce aux outils d'Extract Transform Load (ETL). Le Data Warehouse doit répondre aux critères suivants :

- Orienté métiers : C'est à dire que sa structure est conçue en fonction des besoins utilisateurs. Il se base généralement sur une topologie en étoile et privilégie les performances de réponse au détriment de la normalisation.
- Intègre : Nécessaire pour obtenir des résultats fiables et cohérents.
- Non volatile : Un Data Warehouse est une base en lecture seule.
- Historisé : Toutes les informations sont datées et aucune ne peut être supprimée.

Le Data Warehouse repose sur deux types de tableaux :

1. La dimension est une table qui représente un axe d'analyse suivant lequel on étudie les données observables. Toutefois soumises à une analyse multidimensionnelle, les dimensions donnent aux utilisateurs des renseignements nécessaires à la prise de décision (page 13 du [KR13]).
2. Une table de faits. C'est la table centrale du modèle dimensionnel. Les informations qui s'y trouvent et que nous désirons mesurer doivent être numériques afin d'appliquer des fonctions multi-lignes (somme, moyenne...) (page 10 du [KR13]).

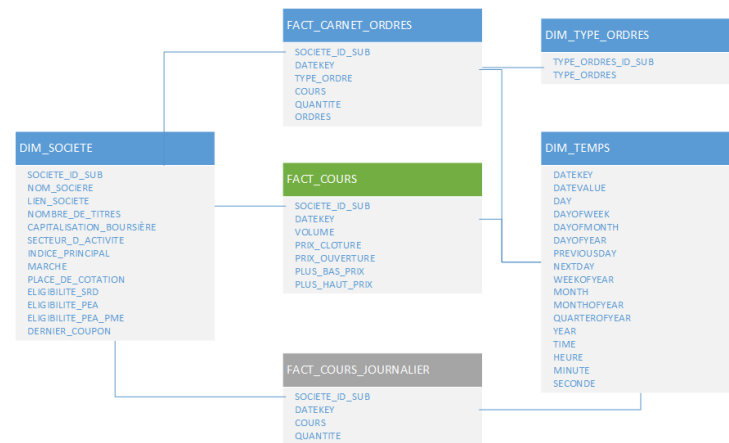


FIGURE A.3. – Modèle du Data Warehouse.

Conception du modèle Data Warehouse Après avoir identifié les axes d'analyse que nous allons mettre en évidence dans les tableaux de bord, nous avons procédé à la conception du Data Warehouse. Il est décrit dans la figure A.3 qui contient 3 dimensions et trois tables de faits.

A.2.3. Chargement du Data Warehouse

Les processus Extraction, Transformation et Chargement (ETL), présenté dans la figure A.4, se chargent de :

- L'extraction des données à partir des bases de données intermédiaires (les données provenant à partir du site web).

A.2. Réalisation de l'application

- La transformation des données extraites (dans l'étape précédente) afin d'éliminer leurs anomalies : les standardiser, les corriger, les normaliser, les rendre cohérents, ...
- Le chargement des données transformées dans le système décisionnel : DataWarehouse, Datamarts.

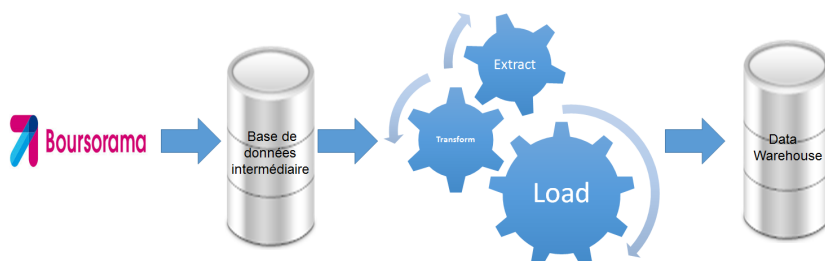


FIGURE A.4. – Processus de l'ETL.

Outil technique utilisé Dans notre projet, nous avons opté pour l'utilisation du Talend Open Studio.

Représentation de Talend

Talend Open Studio est une solution open source, puissante et polyvalente qui sert à l'intégration de données. Elle est conçue pour améliorer la qualité et l'efficacité dans les processus ETL à travers une interface utilisateur graphique. Il s'agit d'un projet Open Source d'intégration de données basé sur Eclipse RCP qui prend en charge les implémentations principalement orientées ETL et est prévu pour le déploiement sur site ainsi que dans un modèle Software-as-a-Service (SaaS) de livraison.

Talend Open Studio est utilisé principalement pour l'intégration entre systèmes opérationnels, ainsi que pour l'ETL (Extract, Transform, Load) pour la Business Intelligence et Data Warehousing, et la migration. Talend propose une vision totalement nouvelle, comme en témoigne la façon dont il utilise la technologie, ainsi que dans son modèle d'affaires. La société brise le modèle propriétaire classique en offrant des solutions logicielles ouvertes, innovantes et performantes avec la souplesse nécessaire pour répondre aux besoins d'intégration de données de tous types d'organisations.

Caractéristiques de Talend

Talend est considéré comme l'un des leaders des outils d'intégration de données grâce à ses caractéristiques. Ceci a été affirmé par l'entreprise Gartner⁷ qui a publié un article qui a procédé à une comparaison des différents outils présents actuellement sur le marché et Talend en a fait preuve d'excellence. Les principales caractéristiques et avantages de cet outil sont :

- La modélisation d'affaires.

- Le développement graphique.
- La conception et l'exécution basée sur les métadonnées.
- Le débogage en temps réel.
- L'exécution robuste.
- La liste variée de composants pour répondre aux besoins des utilisateurs.
- La connexion avec les différents types de sources de données (services web, fichiers Excel, fichiers plats, les différences SGDB, ...).
- La possibilité d'utilisation du langage Java pour faire des traitements personnalisés.
- La communauté est très active.

Création des processus (Job) Talend Dans cette partie, nous allons procéder à l'alimentation du Data Warehouse en effectuant les transformations et les conversions nécessaires pour que les données brutes (provenant du site web) soient conformes aux contraintes du Data Warehouse. Ces transformations se font au niveau des JOBS.

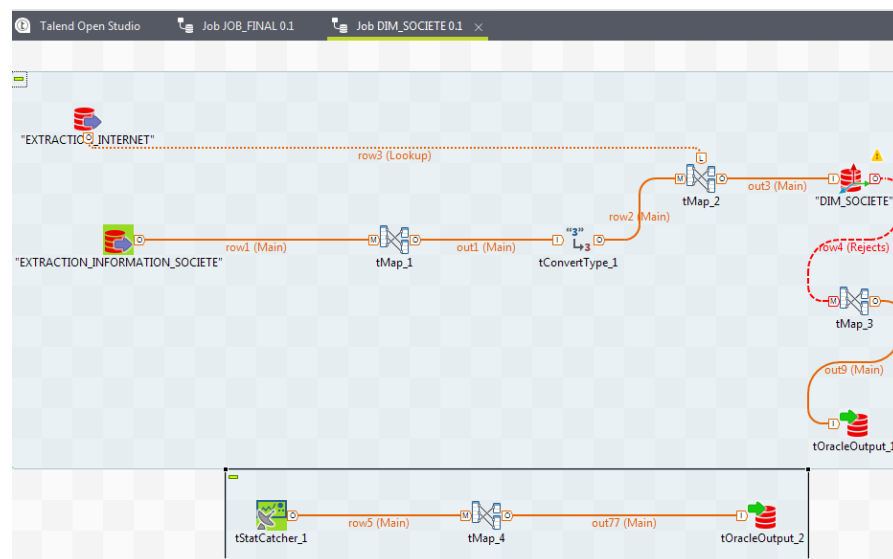


FIGURE A.5. – Chargement de la dimension DIM_SOCIETE.

La figure A.5 désigne le JOB au niveau duquel nous avons effectué les transformations nécessaires afin d'alimenter la dimension DIM_SOCIETE.

Pour alimenter les dimensions, nous avons utilisé le composant « Slowly Changing Dimensions » qui traite les cas d'évolutivité de certains attributs au fil du temps. Dans ce cas, nous avons utilisé SCD de type 1 (qui écrase l'ancienne valeur et la remplace par la nouvelle valeur (à partir de l'historique)) car dans cette dimension nous n'avons pas des attributs qui évoluent au cours du temps.

Nous avons également créé des tables logs pour enregistrer les lignes rejetées afin de faciliter la tâche d'identification des problèmes à l'administrateur des Jobs et une table qui enregistre les diverses informations sur les Jobs (durée d'exécution, le résultat d'exécution des Jobs (Succès, échecs), ...).

A.2. Réalisation de l'application

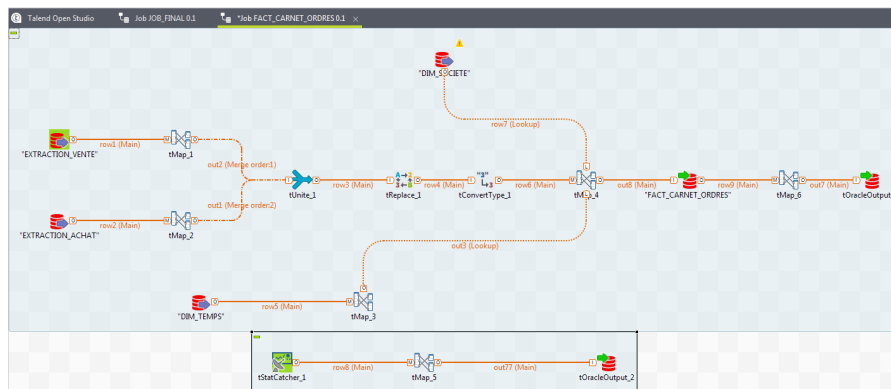


FIGURE A.6. – Chargement de la table de fait FACT_CARNET_ORDER.

La figure A.5 désigne le JOB au niveau duquel nous avons effectué les transformations nécessaires afin d'alimenter la table de fait FACT_CARNET_ORDER et effectué les recherches dans les dimensions afin de récupérer les clefs primaires de ces dernières.

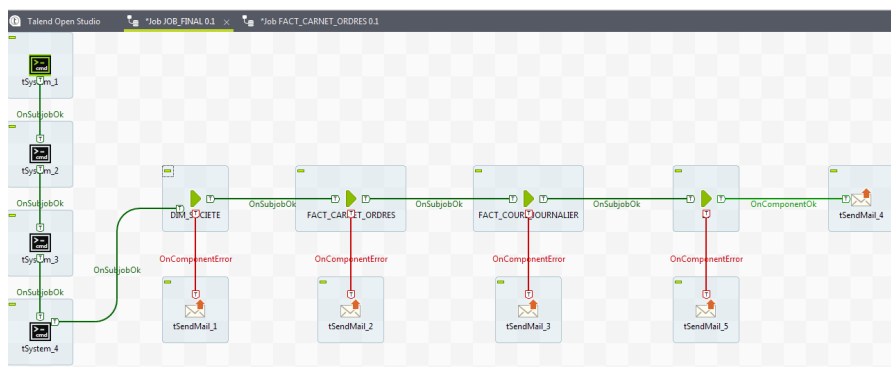


FIGURE A.7. – Exécution de tous les JOBS.

La figure A.7 désigne le JOB au niveau duquel nous avons invoqué les différentes classes Java afin d'extraire les données à partir du web, les insérer dans les bases de données intermédiaires. Ensuite, nous procédons à l'alimentation du Data Warehouse. Si le JOB s'exécute normalement, l'administrateur de l'ETL reçoit un mail qui confirme le bon déroulement du JOB sinon il reçoit un mail d'échec.

A.2.4. Réalisation des tableaux de bord

Après avoir alimenté le Data Warehouse, nous procédons à la création des tableaux de bords interactifs pour mettre en évidence les indicateurs et les axes d'analyse présélectionnés comme l'indique la figure A.8.

Ces tableaux de bord seront mis à jour d'une façon automatique et régulière à travers les processus ETL réalisés auparavant.

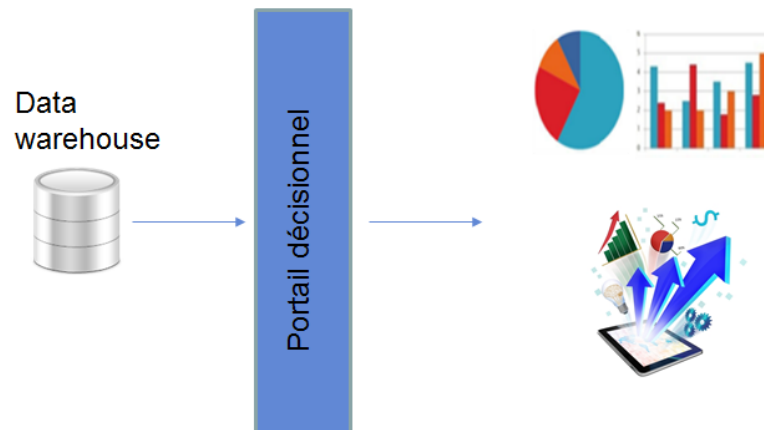


FIGURE A.8. – Processus de restitution des données.

Outil technique utilisé Dans notre projet, nous avons utilisé Tableau Desktop. Ceci est une solution polyvalente pour la restitution des données, l'analyse de ces dernières et la réalisation des tableaux de bords, qui sont conçus pour améliorer la capacité d'exploration des données à travers une interface utilisateur graphique.

Représentation de Tableau

Une recherche sur l'analyse et la visualisation des données « VizQL » a été menée en 2003 à l'université de Stanford et a été l'objet de création de l'éditeur Tableau Software. Pour plus d'informations, le lecteur peut se référer à l'article [MS].



FIGURE A.9. – Logo Tableau Software.

Caractéristiques de Tableau

Les principales caractéristiques et avantages de cette solution sont :

- La facilité de manipulation des données.
- La communication avec un serveur R pour réaliser les diverses études data mining.
- La possibilité d'exporter les tableaux de bord sur les diverses plateformes mobiles (Android, IOS) et optimisation automatique de la taille de ces derniers sur les différentes tailles d'écrans.

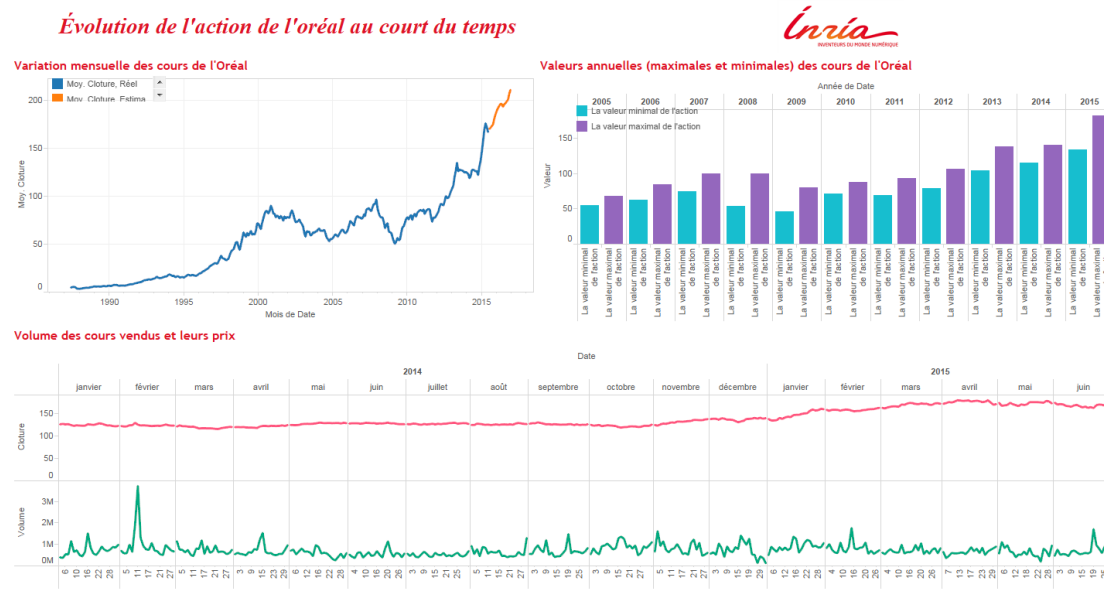


FIGURE A.10. – Tableau de bord du L'Oréal.

- La possibilité de fusionner des données hétérogènes en un seul clic (relier plusieurs sources de données (Bases de données, fichiers Excel ...)).
- La rapidité et la robustesse de manipulation des données de gros volume.

Réalisation des Tableaux de bord

Pour réussir la mise en place des tableaux de bord, il faut penser à mettre en évidence les principaux axes d'analyse qui influent essentiellement le processus de prises de décisions. Nous soulignons que notre étude doit être ciblée et donc nous allons éviter de charger les tableaux de bords avec une multitude d'indicateurs non significatifs.

Ce tableau de bord illustre l'évolution du prix de l'action du L'Oréal au fil des années. Il représente également le volume des cours vendus en fonction du prix. Nous avons aussi effectué la prévision de l'évolution du prix des cours de cette action sur deux années. Pour ce faire, nous nous sommes basé sur l'algorithme du lissage exponentiel qui est intégré dans notre outil.

Lissage exponentiel C'est un outil de modélisation de prévision des séries temporelles. Son principe est de donner plus d'importance aux dernières observations. Cette technique modélise les prévisions ultérieures d'une manière redondante d'une série temporelle régulière. Cela revient à donner un poids important aux données récentes et de faire décroître ce poids d'une manière proportionnelle en traitant des données plus anciennes. C'est le cas général de la EWMA qui est utilisée pour le calcul de la variance du prix de l'actif.

Le Tableau Software recherche un cycle saisonnier (un acte redondant sur une période de temps) avec la longueur la plus typique de l'agrégation chronologique de la série

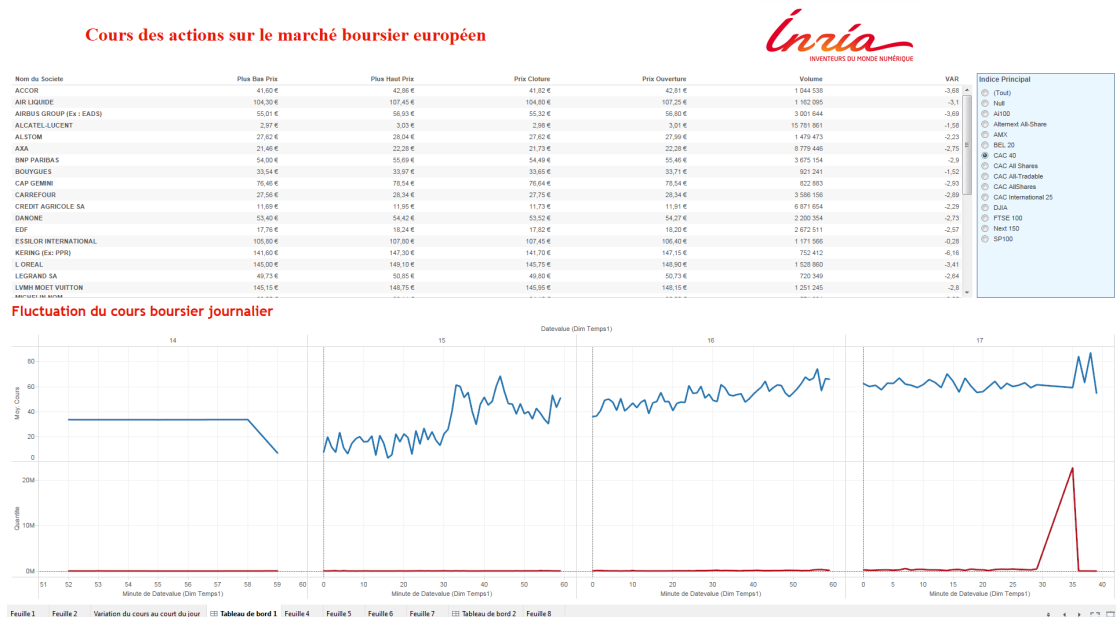


FIGURE A.11. — Tableau de bord du marché financier européen.

temporelle. Pour plus de détails, le lecteur peut se référer au livre [Ara11].

Le tableau de bord de la figure A.11 illustre l'évolution du marché boursier européen au cours d'une journée donnée. Nous pouvons choisir l'indice boursier selon lequel nous souhaitons effectuer notre analyse financière. Nous y trouvons également la *VaR* de chaque actif.

Déploiement des tableaux de bord sur le portail décisionnel et les appareils mobiles

Nous avons pensé à exporter nos tableaux de bord sur des appareils mobiles afin que les décideurs arrivent à y accéder et obtenir les informations nécessaires. Cela leur permet d'avoir une idée sur l'évolution du marché boursier. La solution mobile de Tableau fournit :

1. La possibilité de consulter les tableaux de bord à tout moment.
2. Des vues optimisées pour les écrans tactiles.
3. La possibilité de consulter les tableaux de bord à tout moment.
4. Une expérience cohérente d'un périphérique à l'autre.

Nous avons déployé les tableaux de bords créés auparavant sur le serveur Tableau Online (la figure A.13) pour pouvoir bénéficier des fonctionnalités de Tableau Server tels que l'accès aux tableaux de bords à travers les appareils mobiles ou les navigateurs web. Le serveur Tableau Online permet d'assurer l'interaction entre les décideurs et ceci grâce à la centralisation de l'accès aux tableaux de bord. Les vues de Tableau sont optimisées de manière à offrir des expériences tactiles lorsqu'un utilisateur y accède sur un iPad ou sur une tablette Android (la figure A.12). La compatibilité avec les écrans tactiles est automatique : il n'est pas nécessaire de procéder à des modifications en termes de création ou de conception pour que les vues et les tableaux de bord fonctionnent automatiquement

A.3. Conclusion

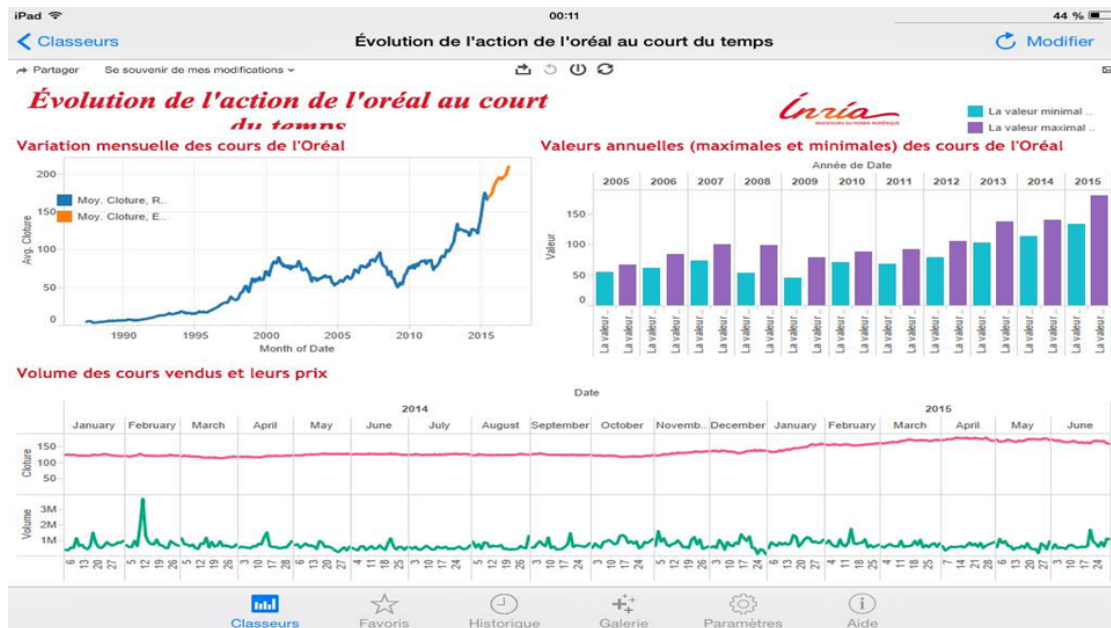


FIGURE A.12. – Tableau de bord de L'Oréal sur la tablette.

sur des tablettes.

A.3. Conclusion

L'étude approfondie du processus métier de l'évolution du marché financier nous a aidé à identifier les principaux axes d'analyse. Les tableaux de bord réalisés nous ont conduit à extraire des informations pour formuler des conclusions afin d'optimiser les décisions et minimiser le risque d'investisseurs.

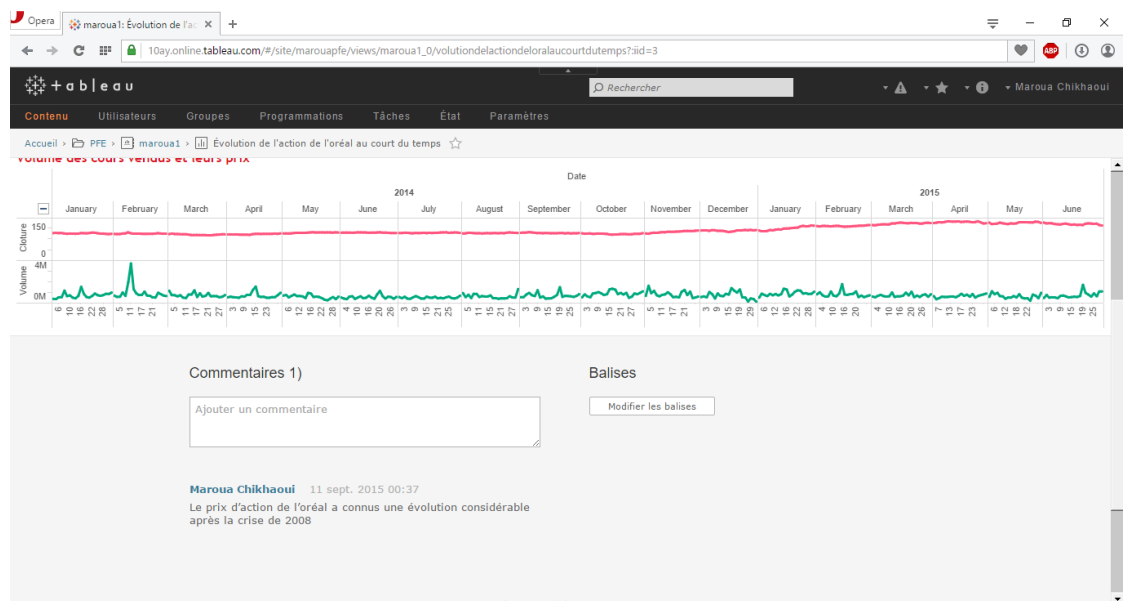


FIGURE A.13. – Tableau de bord du marché financier européen en ligne.

Bibliographie

- [Ara11] Yves Aragon. *Séries temporelles avec R—méthodes et cas*. Pratique R. Springer, Paris, 2011. With a preface by Dominique Haughton.
- [AT02a] Carlo Acerbi and Dirk Tasche. Expected shortfall : a natural coherent alternative to value at risk. *Economic Notes*, 31(2) :379–382, 2002.
- [AT02b] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7) :1487 – 1503, 2002.
- [BW12] Jaroslav Baran and Jiří Witzany. A comparison of EVT and standard VaR estimations. *Bulletin of the Czech Econometric Society*, 19(29), 2012.
- [KR13] Ralph Kimball and Margy Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*. Wiley Publishing, 3rd edition, 2013.
- [Lek10] Alexandre Lekina. *Nonparametric estimation of conditional extreme quantiles*. Thèses, Université Joseph-Fourier - Grenoble I; Université de Grenoble, October 2010.
- [LL92] Yoong-Sin Lee and Ting-Kwong Lin. Statistical algorithms : Algorithm AS 269 : High order Cornish-Fisher expansion. *Journal of Applied Statistics*, 41(1) :233–240, 1992.
- [Lon94] Jacques Longestaey. Riskmetrics technical document. Technical report, JP Morgan, 1994. Available at : <http://yats.free.fr/papers/td4e.pdf>.
- [LP02] Jye-Chyi Lu and Liang Peng. Likelihood based confidence intervals for the tail index. *Extremes*, 5(4) :337–352, 2002.
- [MS] Pat Hanrahan Jock D. Mackinlay and Chris Stolte. Automatic presentation for visual analysis.
- [PS99] Stefan Pichler and Karl Selitsch. A comparison of analytical VaR methodologies for portfolios that include options, 1999. Working Paper TU.